ERRATA

THE ANNALS OF MATHEMATICAL STATISTICS

Volume VI, No. 1, March, 1935

On page 25, in Directions for Use of the Tables, $p = q$ should read $p \gtrless q$, $Q_1 = x_1 + qn$ should read $Q_1 = x_1\sigma + qn$, $D_9 = x_2 + qn$ should read $D_9 = x_2\sigma + qn$. In the tables of values of x under $p = .97$, $n = 25$, instead of $-.784$ the number should be $-.754$.

AN APPLICATION OF ORTHOGONALIZATION PROCESS TO THE THEORY OF LEAST SQUARES

By Y. K. Wong

Introduction

The present paper is an outgrowth of the writer's attempt to fill a lacuna in the discussion of the Gauss method of substitution as given by many writers. For illustration, let us cite Brunt's *Combination of Observations*. In Chapter VI, we find:

Let the normal equations be

$$\begin{aligned} [aa]x + [ab]y + [ac]z - [al] &= 0 \\ [bb]y + [bc]z - [bl] &= 0 \\ [cc]z - [cl] &= 0 . \end{aligned} \tag{i}$$

From this equation we find

$$x = -\frac{[ab]}{[aa]}y - \frac{[ac]}{[aa]}z + \frac{[al]}{[aa]} . \tag{ii}$$

Substituting, we obtain

$$\begin{aligned} [bb1]y + [bc1]z - [bl1] &= 0 \\ [cc1]z - [cl1] &= 0 \end{aligned} \tag{iii}$$

where

$$[bb1] = [bb] - [ab]\,[ab]/[aa], \text{ etc.} \tag{iv}$$

From the first equation in (iii),

$$y = -\frac{[bc1]}{[bb1]}z + \frac{[bl1]}{[bb1]} . \tag{v}$$

In connection with equations (ii) and (v), the question naturally arises as to whether or not these numbers $[aa]$, $[bb1]$, $\cdots$ are all different from zero. Since $[aa] = \Sigma a_i a_i$, one can see that $[aa] \neq 0$ if $a_i \neq 0$ for every i. However, to show the non-vanishing of $[bb.1]$, $[cc.2]$, etc. is by no means simple. Many writers do not give a demonstration on this point. We know that a system of non-homogeneous linear equations has a solution if the system of equations is linearly independent. Brunt gives a discussion of the independence of the normal equations in Chapter V, Art. 36, but he does not state clearly a condition for independence. He says: "The condition of independence is in general satisfied in

the problems which arise in practice. We can then proceed to the formation and solution of the normal equations." It is one of the aims of this paper to give a necessary and sufficient condition for the independence of the normal equations and to show $[aa]$, $[bb.1]$, etc. are all different from zero when the condition is satisfied.

In the theory of least squares, there is the classical method of the derivation of normal equations by an application of the notion of minimum in differential calculus. After the normal equations are secured, the Gauss method of substitution is applied to obtain the solution. Doolittle modifies the Gauss method of substitution so as to facilitate the labor of computation. However, when the number of parameters (or unknowns) exceeds 4, Doolittle's method is quite complicated. In the present paper the writer wishes to present a mathematical discussion of a method obtained through an application of the Gram-Schmidt orthogonalization process. This method furnishes us a new procedure for determining the most probable values of the parameters (or unknowns). The formulation of the system of normal equations will be omitted in this new procedure, which is particularly effective in fitting curves to time series. The paper can be roughly divided into three parts. The first part gives an algebraic derivation of the normal equations. The second part derives a condition for a set of observation data so that the Gauss method of substitution is applicable. The third part gives a relationship between the Gauss method of substitution and the orthogonalization process. A practical application of the results of this paper will be found in a later paper.

The process of orthogonalization has been used in the 19th century, and has been applied extensively in the theory of integral equations and linear transformations in Hilbert space. In classical analysis, if $\varphi_1(x), \varphi_2(x), \cdots$, defined on (0, 1), is a normally orthogonalized system, and if $f(x)$, defined on (0, 1), is such that f^2 is Lebesgue integrable, then the system of Fourier coefficients

$$f_r = \int_0^1 f(x)\varphi_r(x)dx \qquad (r = 1, 2, \cdots)$$

has certain interesting properties, one of which is that

$$\frac{1}{m}\int_0^1 (f(x) - \sum_1^m f_r\varphi_r)^2\, dx = 0\,.$$

The preceding notion has a close connection with the theory of least squares as outlined in many texts on statistics. In section III, the reader will find how this notion is applied in the derivation of the normal equations. Since the number of dimensions is finite, the integration process reduces to a summation process and furthermore no limiting process is used. This new derivation of normal equations has the advantage that (1) differential calculus is not used, (2) a new form of normal equations is obtained, (3) the solution of the unknowns or parameters can be immediately obtained without further application of the

Gauss Method of Substitution or the Doolittle Method, and (4) the formula for the "quadratic residual" is obtained as a simple corollary.

From the results in section III, we see immediately what condition should be imposed upon the set of observation data so that the Gauss method of substitution may be applicable. In section VI, we find a necessary and sufficient condition for the independence of the system of normal equations (3.9), and also the fact that when this condition is fulfilled, then, due to the special nature of the coefficients of the unknowns, we see that the matrix is properly positive. It is on account of this fact that we are able to show that the numbers $[aa]$, $[bb.1]$, etc. are all different from zero. The demonstration of this point is found in section VII. In this section, we lay down a fundamental hypothesis for Gauss's method of substitution, namely, the set of observations $A_i = (a_{i1}, \cdots, a_{in})$ $i = 1, 2, \cdots, r$, is linearly independent. Lemma 7.3 may be called the fundamental lemma for Gauss's method of substitution. Some interesting properties of the numbers $[A_sA_t \cdot h]$, where $s, t = 1, \cdots, r$, and h is less than the smaller one of (s, t), are demonstrated.

From the properties of the numbers $[A_sA_t \cdot h]$, where $s, t = 1, \cdots, r$ and h is less than the smaller one of (s, t), and in comparison of the system of equations (3.7^0) with the final form of equations obtained through the application of the Gauss method of substitution, we can see the relationship between the Gauss method and the Gram-Schmidt orthogonalization process. If we should like to give some credit to Gauss, we may say that the orthogonalization process was known by him, but was stated in a different form.

The writer wishes to remark that certain theorems together with proofs in section II, IV, V and VI are obtained from E. H. Moore's lecture notes. However the writer should be responsible for any defect. Finally, I should emphasize that the use of the notion of positive matrices is only for convenience.

I. Vectors, Inner Products, and Linear Independence

In this paper, we shall consider vectors of the form[1]

$$(v_1, v_2, \cdots, v_n). \tag{1.10}$$

For convenience, we shall use capital letters to denote vectors of the type (1.10).

Let $V = (v_1, v_2, \cdots, v_n)$ and $U = (u_1, u_2, \cdots, u_n)$, then we say $V = U$ if $v_i = u_i$ for every i.

We define $V + U$ by

$$V + U = (v_1 + u_1, v_2 + u_2, \cdots, v_n + u_n), \tag{1.11}$$

and sV, where s is a number, by

$$sV = (sv_1, sv_2, \cdots, sv_n). \tag{1.12}$$

[1] If we write v_i as $v(i)$, where $i = 1, 2, \cdots, n$, then $v(i)$ may be considered as a function of one variable whose range consists of a set of positive integers, $(1, 2, \cdots, n)$. E. H. Moore defines a vector as a function of one variable.

Hence, $sV = Vs$. In particular, when $s = -1$. we shall put $-V = (-1)V$. Then $U - V$ becomes a special instance of (1.11) and (1.12).

From (1.11) and (1.12), we see that addition is commutative and associative.

INNER PRODUCTS: The inner product of two vectors $V = (v_1, \cdots, v_n)$ and $U = (u_1, \cdots, u_n)$ is defined[2] to be

$$(V, U) = \sum_1^n v_i u_i . \tag{1.2}$$

The norm of a vector V is defined by $n(V) = (V, V)$; and the modulus of a vector V is defined by mod $(V) = + \sqrt{n(V)}$.

From (1.11), (1.12), and (1.2), we can easily prove the following theorem:

THEOREM 1. *The symbol* (,) *has the following properties:*

(S) $(U, V) = (V, U)$ *for every* V, U; (symmetric property)

(L_s) $(sV, U) = s(V, U) = (V, sU)$ *for every* V, U *and every number* s;

(L_+) $(U, (V + W)) = (U, V) + (U, W)$ *for every* U, V, W; (*linear property*)

(P) $(V, V) \geqq 0$ *for every* V; (*positive property*)

(P_0) $(V, V) = 0$ *if and only if* V *is a zero vector*; (*properly positive property*)

LINEAR INDEPENDENCE. A set of vectors $V_1, \cdots, V_r$ is said to be linearly dependent in case there exist constants $c_1, \cdots, c_r$ not all equal to 0 such that

$$c_1 V_1 + \cdots + c_r V_r = 0 ,$$

where 0 is a zero vector.

A set of vectors $V_1, \cdots, V_r$ is said to be linearly independent in case, if the constants $c_1, \cdots, c_r$ satisfy

$$c_1 V_1 + \cdots + c_r V_r = 0 ,$$

each constant $c_i = 0$.

THEOREM 2. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then none of the vectors is a zero vector, and hence the norm of every vector must be different from zero.*

For if V_s is a zero vector, then set $c_s = 1$, and $c_i = 0$ for $i \neq s$. It is obvious that

$$0 \cdot V_1 + \cdots + 0 \cdot V_{s-1} + 1 \cdot V_s + 0 \cdot V_{s+1} + \cdots + 0 \cdot V_r = 0 ,$$

which show that the set of vectors $V_1, \cdots, V_r$ is linearly dependent, contradictory to the hypothesis.

A more general theorem is stated in

THEOREM 3. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then every subset*[3] *is also linearly independent.*

We shall prove this theorem by a contrapositive form. The contrapositive form is as follows: *If in the set* $V_1, \cdots, V_r$, *there exists a subset which is linearly*

[2] The notation (,) was introduced by D. Hilbert. In treatises on least squares, the notation [ ] is used. The present writer reserves the latter notation for other purposes.

[3] Consider a set of integers $(1, 2, \cdots, n)$. Then any *combination* of this set of n distinct integers taken $r \leqq n$ at a time is called a *subset* of the set $(1, 2, \cdots, n)$. Likewise, we call any combination of the set of vectors $V_1, V_2, \cdots, V_n$ taken $r \leqq n$ at a time a *subset* of the whole set.

dependent, then the whole set is also linearly dependent. Without losing any generality, let us suppose the subset $V_1, \cdots, V_s$ $(s \leqq r)$ to be linearly dependent. Then there exist $c_1, \cdots, c_s$ such that

$$c_1V_1 + \cdots + c_sV_s = 0\,.$$

If $s = r$, then the whole set is linearly dependent. If $s < r$, then let $c_i = 0$ for $i = s - 1, s - 2, \cdots, r$. Then

$$\sum_1^r c_iV_i = 0\,,$$

which shows the whole set is linearly dependent.

THEOREM 4.[4] *A necessary and sufficient condition for the set* $V_i = (v_{i1}, \cdots, v_{in})$, $i = 1, \cdots, r$ *to be linearly independent is that there exists a non-vanishing determinant of order* r *in the array*

$$\begin{matrix} v_{11}, v_{12}, \cdots, v_{1n} \\ v_{21}, v_{22}, \cdots, v_{2n} \\ \cdots\cdots\cdots \\ v_{r1}, v_{r2}, \cdots, v_{rn} \end{matrix}\,.$$

II. Gram-Schmidt's Orthogonalization Process

For the present section and the sequel, we shall adopt the notation $A_i = (a_{i1}, \cdots, a_{in})$, $B_i = (b_{i1}, \cdots, b_{in})$, and $C_i = (c_{i1}, \cdots, c_{in})$ for $i = 1, 2, \cdots, r$.

THEOREM 5. *For every set of vectors* $A_1, \cdots, A_r$, *there exists uniquely a set of vectors* $B_1, \cdots, B_r$ *such that*

5.1) $(B_t, B_s) = 0$ $(t \neq s)$.

5.2) *For every* t *satisfying the relation* $1 \leqq t \leqq r$, *then* A_t *is a linear combination of* $B_1, \cdots, B_t$; *and* B_t *is a linear combination of* $A_1, \cdots, A_t$.

5.3) $B_1 = A_1$; *and for* $t > 1$, $(B_t - A_t)$ *is a linear combination of* $B_1, \cdots, B_{t-1}$, *and is also a linear combination of* $A_1, \cdots, A_{t-1}$.

5.4) *If* $t > 1$, *then* $(A_s, B_t) = 0$ *for every* $s < t$.

5.5) $(A_t, B_t) = (B_t, B_t) = (B_t, A_t)$ *for every* t.

To prove this theorem, let us define

$$\begin{aligned} B_1 &= A_1\,, \\ B_2 &= A_2 \quad \text{if} \quad n(B_1) = 0 \\ &= A_2 - \frac{(A_2, B_1)}{n(B_1)} B_1 \quad \text{if} \quad n(B_1) \neq 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ B_t &= A_t - \sum_{i=1}^{t-1} h_{ti}B_i \qquad (1 \leqq t \leqq r)\,, \end{aligned} \tag{2.1}$$

[4] See Dickson, *Modern Algebraic Theories*, p. 55; Bocher, *Higher Algebra*, p. 36.

where

$$h_{ti} = (A_t, B_i)/n(B_i)\,, \qquad \text{if } n(B_i) \neq 0\,,$$
$$= 0\,, \qquad \text{if } n(B_i) = 0\,. \tag{2.11}$$

We proceed to show that this set has the properties stated in the theorem.

To prove 5.1), let us suppose $t < s$. This assumption is permissible since the operator $(\ ,\ )$ has the symmetric property. First, if $A_1 = 0$, then $B_1 = 0$, and

$$(B_1, B_2) = (A_1, A_2) = (0, A_2) = 0\,.$$

Secondly, if $A_1 \neq 0$, then $B_1 \neq 0$ and

$$(B_1, B_2) = (A_1, A_2 - h_2, B_1) = (A_1, A_2) - (A_1, B_1)\frac{(A_2, B_1)}{n(B_1)}$$
$$= (A_1, A_2) - (A_1, A_1)\,(A_2, A_1)/n(A_1) = 0\,.$$

Assume 5.1) is true for $t = s - 1$, then

$$(B_t, B_s) = \left(B_t, A_s - \sum_1^{s-1} h_{si}B_i\right) = (B_t, A_s) - \sum_1^{s-1} h_{si}(B_t, B_i)\,.$$

The sum on the right hand side reduces to $h_{st}(B_t, B_t)$, since the other terms vanish by assumption. Now if $(B_t, B_t) \neq 0$ then by (2.11), $h_{st}(B_t, B_t) = (A_s, B_t)$, and by the symmetric property of $(\ ,\ )$, we obtain

$$(B_t, B_s) = (B_t, A_s) - (A_s, B_t) = 0\,.$$

If $(B_t, B_t) = 0$, then by the P_0-property of $(\ ,\ )$, we find that B_t is a zero vector, and hence $(B_t, B_s) = 0$.

5.2) follows from the definition of B_t.

That $(A_t - B_t)$ is a linear combination of $B_1, \cdots, B_{t-1}$ for $t > 1$ follows from the definition of B_t. Since B_s is a linear combination of $(A_1, \cdots, A_{s-1})$, we secure the second part of 5.3).

By 5.2), we can determine g_{si} such that $A_s = \sum_1^s g_{si}B_i$. Thus for every $s < t$, we have by 5.1)

$$(A_s, B_t) = \left(\sum_1^s g_{si}B_i, B_t\right) = \sum_1^s g_{si}(B_i, B_t) = .0$$

By 5.3), there exist g_{ti} such that $A_t - B_t = \sum_1^{t-1} g_{ti}B_i$ and hence $A_t = B_t + \sum_1^{t-1} g_{it}B_i$. Thus by 5.1), we have

$$(A_t, B_t) = \left(B_t + \sum_1^{t-1} g_{ti}B_i, B_t\right) = (B_t, B_t) + \sum_1^{t-1} g_{ti}(B_i, B_t)$$
$$= (B_t, B_t)\,.$$

By the symmetric property of $(\ ,\ )$, we secure $(A_t, B_t) = (B_t, B_t)$.

For the proof of uniqueness, let us suppose there exists a second set of vectors $B_1', \cdots, B_r'$ having the properties 5.1), 5.2), 5.3), 5.4), and 5.5). By 5.3), we see that $B_1 = A_1 = B_1'$. Assuming the uniqueness holds true for $r = t$, we proceed to show that it is also true for $r = t + 1$. By 5.3) there exist constants s_i, s_i' $(i = 1, \cdots, t)$ such that

$$B_{t+1} = A_{t+1} + \sum_1^t s_i A_i$$

$$B'_{t+1} = A_{t+1} + \sum_1^t s_i' A_i .$$

Thus

$$B_{t+1} - B'_{t+1} = \sum_1^t (s_i - s_i') A_i .$$

From this, we secure

$$(B_{t+1} - B'_{t+1}, B_{t+1} - B'_{t+1}) = \left(B_{t+1} - B'_{t+1}, \sum_1^t (s_i - s_i') A_i\right)$$

$$= \sum_1^t (s_i - s_i') \cdot (B_{t+1} - B'_{t+1}, A_i) = 0 ,$$

by virtue of 5.4). Hence by P_0-property of (,), we have $B_{t+1} - B'_{t+1} = 0$ and hence $B_{t+1} = B'_{t+1}$.

The set $B_1, \cdots, B_r$ with the properties stated in Theorem 5 is called the *orthogonalized set* of $A_1, \cdots, A_r$. This process is called Gram-Schmidt's orthogonalization process.

The set $B_1, \cdots, B_r$ is called the *normally orthogonalized set* of $A_1, \cdots, A_r$ in case the former set enjoys the properties 5.1), 5.2), 5.3), 5.4), and if

5.5n) $\qquad (A_t, B_t) = (B_t, B_t) = (B_t, A_t) = 1$ *for every* t .

THEOREM 6. *If a subset* $A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$ *in the set* $A_1, \cdots, A_r$, *is linearly independent, then there is a subset* $B_{k_1}, \cdots, B_{k_m}$ *which has the properties stated in Theorem* 5, *and it is also linearly independent.*

Let $h = k_m - k_1 + 1$. To prove the theorem, we may assume $k_1, \cdots, k_m$ to be $1, \cdots, h \leqq r$, for otherwise, we may renumber the vectors. We construct the B vectors in the same way as given in equation (2.1) and (2.11). By Theorem 5, we have

(2.2) $\qquad B_1 = A_1, \qquad B_s = A_s + \sum_1^{s-1} g_{si} A_i \qquad (s = 2, \cdots, h) .$

Suppose the constants $c_1, \cdots, c_h$ be such that

$$c_1 B_1 + \cdots + c_h B_h = 0 .$$

Then by (2.2), we secure

$$0 = c_1A_1 + \sum_2^h c_sB_s = c_1A_1 + \sum_2^h c_s\left(A_s + \sum_1^{s-1} g_{si}A_i\right)$$

$$= (c_1 + c_2g_{21}, + \cdots + c_hg_{h1})\,A_1 + (c_2 + c_3g_{32} + \cdots c_hg_{h2})A_2 + \cdots + c_hA_h\,.$$

Since $A_1, \cdots, A_h$ are linearly independent, we have

$$\begin{aligned} c_1 - c_2g_{21} - \cdots - c_hg_{h1} &= 0\,, \\ c_2 - \cdots - c_hg_{h2} &= 0\,, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \\ c_h &= 0\,. \end{aligned} \tag{2.3}$$

But the determinant of the coefficients of $c_i (i = 1, \cdots, h)$ is

$$\begin{vmatrix} 1 & g_{21} & g_{31} & \cdots & g_{h1} \\ 0 & 1 & g_{32} & \cdots & g_{h2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} = 1\,.$$

Hence by a theorem in the theory of equations,[5] the only solution that satisfies (2.3) is that $k_1 = k_2 = \cdots = k_h = 0$. Thus the subset $B_1, \cdots, B_h$ is linearly independent.

COROLLARY. *The orthogonalized set $B_1, \cdots, B_r$ is linearly independent if and only if the set $A_1, \cdots, A_r$ is linearly independent.*

THEOREM 7a. *If a set of vectors $A_1, \cdots, A_r$ is linearly independent, then the set can be normally orthogonalized.*

Let B_i be the orthogonalized set of A_i. Since A_i is a linearly independent set, then the set B_i is also linearly independent by Theorem 6. Hence by Theorem 2, the norm of every vector B_i is non-vanishing. Define $C_i = B_i/\text{mod}\ (B_i)$. Then this set C_i enjoys the properties 5.1), 5.2), 5.3), 5.4) and 5.5n).

THEOREM 7b. *If a set of vectors, $V_1, \cdots, V_r$ is normally orthogonal, i.e.* if

$$(V_i, V_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j)\,, \end{cases} \tag{2.4}$$

then $V_1, \cdots, V_r$ is linearly independent.

For suppose

$$c_1V_1 + \cdots + c_rV_r = 0\,.$$

Then

$$\sum_{i=1}^{r} c_i(V_i, V_j) = 0\,, \qquad (j = 1, 2, \cdots, r)\,.$$

[5] Dickson, *First Course in the Theory of Equations* (1922), p. 119.

By condition (2.4), the preceding expression reduces to

$$c_j = 0, \qquad (j = 1, 2, \cdots, r),$$

which shows the linear independence of $V_1, \cdots, V_r$.

III. Algebraic Derivation of the Normal Equations

Consider a linear function

$$l = p_1x_1 + p_2x_2 + \cdots + p_rx_r = \sum_1^r p_ix_i. \tag{3.1}$$

Let the set of observations of x_i and l be

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots, r\,;\, n \geqq r) \tag{3.2}$$

respectively, then the residual v_i is

$$v_i = \sum_{j=1}^r p_ja_{ji} - l_i, \qquad (i, = 1, \cdots, n).$$

In vector notation,

$$V = \sum_{j=1}^r p_jA_j - L.$$

The theory of least squares requires us to find the values for $p_1, \cdots, p_r$ so as to make (V, V) a minimum, or

$$(\sum p_jA_j - L, \sum p_jA_j - L) = \text{a minimum.} \tag{3.3°}$$

Let $A_1, \cdots, A_r$ be linearly independent. By Theorem 7, the vectors $A_1, \cdots, A_r$ can be normally orthogonalized. Let $C_1, \cdots, C_r$ be the normally orthogonal set. Then every A_t $(t = 1, \cdots, r)$ is expressible as a linear combination of $C_1, \cdots, C_t$. Let us write

$$\sum_1^r p_jA_j = \sum_1^r k_jC_j. \tag{3.3}$$

Our problem now is equivalent to that of finding the values $k_i (i = 1, \cdots, r)$ so as to render the inner product

$$(\sum k_jC_j - L, \sum k_jC_j - L) \tag{3.4}$$

a minimum. Expression (3.4) can be written in the form

$$\begin{aligned} &(L, L) - 2\sum(L, C_i)k_i + \sum_{i,j}(k_iC_i, k_jC_j) \\ &= (L, L) - 2\sum(L, C_i)k_i + \sum k_i^2 \qquad (3.5) \\ &= (L, L) - \sum(L, C_i)^2 + \sum(k_i - (C_i, L))^2. \end{aligned}$$

Hence (3.4) gives a minimum if and only if the last summation vanishes, i.e.,

$$k_i = (C_i, L) \qquad (i = 1, \cdots, r). \tag{3.6}$$

The Bessel's inequality

$$\sum_1^r k_i^2 \leqq (L, L)$$

is obtained from (3.6), (3.4), and (3.5).

To solve for p_i, we make use of (3.3) and (3.6), and secure

$$\sum_1^r A_i p_i = \sum_1^r (C_i, L) C_i ,$$

whence

$$\left(C_k, \sum_1^r A_j p_j\right) = \left(C_k, \sum_1^r (C_i, L)\, C_i\right).$$

On the right hand side we have

$$(C_k, \sum (C_i, L) C_i) = \sum (C_i, L)\, (C_k, C_i) = (C_k, L) ,$$

since $(C_k, C_i) = 0$ when $i \neq k$, and $(C_k, C_i) = 1$ when $i = k$. On the left hand side, we have

$$\left(C_k, \sum_{j=1}^r A_j p_j\right) = \sum_{j=1}^r (C_k, A_j) p_j = \sum_{j=k}^r (C_k, A_j) p_j ,$$

since $(C_k, A_j) = 0$ when $j < k$. Hence the values for $p_1, \cdots, p_r$ are given by

$$\sum_{i=k}^r (C_k, A_i) p_i = (C_k, L) \qquad (k = 1, \cdots, r) , \tag{3.7}$$

where $(C_i, A_i) = (C_i, C_i) = 1$.

Equations (3.7) are called the normal equations, which are derived without using any notion in differential calculus.

From (3.6) and (3.5), we secure the value for the 'quadratic residual' (V, V):

$$(V, V) = (L, L) - \sum_{i=1}^r (L, C_i)^2, \tag{3.8}$$

which is a positive quantity by virtue of the Bessel's inequality.

Let $B_1, \cdots, B_r$ be an orthogonalized set of $A_1, \cdots, A_r$. Then every vector B_i has a non-vanishing norm, and $B_i = \text{mod } (B_i) \cdot C_i$. Hence from (3.7) and (3.8), we have

$$\sum_{i=k}^r (B_k, A_i) p_i = (B_k, L), \qquad (k = 1, 2, \cdots, r) , \tag{3.7°}$$

$$(V, V) = (L, L) - \sum_{i=1}^r (L, B_i)^2 / n(B_i) . \tag{3.8°}$$

Thus we have proved the following

THEOREM 8. *Given a linear function* (3.1). *Let the set of observations of* x_i *and* l *be*

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots r; n \geqq r)$$

respectively. Let $A_1, \cdots, A_r$ *be linearly independent,* $B_1, \cdots, B_r$ *be the orthogonalized set, and* $C_1, \cdots, C_r$, *the normally orthogonalized set of* $A_1, \cdots, A_r$. *Then the set of values* $p_1, \cdots, p_r$ *will minimize* (3.3°) *if and only if the system of equations* (3.7°) *or* (3.7) *holds true; in other words,* $\sum_{i=1}^{r} p_i A_i - L$ *is orthogonal to* C_j *or to* B_j *for every* j. *The quadratic residual* (V, V) *is given by* (3.8°) *or* (3.8).

From (3.7), we can secure the solution for $p_1, \cdots, p_r$ immediately without further application of the Gauss method of substitution.

The proof of the following theorem does not make use of the orthogonalization process.[6]

THEOREM 8°. *Let* $F = \Sigma\, p_i A_i$, *where every* A_i *is not a zero vector. The set of values* $p_1, \cdots, p_r$ *will minimize* (3.3°) *if and only if* $(F - L, A_i) = 0$ *for every* i, i.e., $F - L$ *is orthogonal to every* A_i.

The condition is necessary. To prove this, we show that if $(F - L, A_i) \neq 0$ for every i, then we can find another set $q_1, \cdots, q_r$ such that $n(F - L) > n(G - L)$, where $G = \Sigma\, q_i A_i$. For if $(F - L, A_i) \neq 0$ for every i, then we can find a vector A_s such that $(F - L, A_s) \neq 0$. Since $A_s \neq 0$, we let $e = (F - L, A_s)/n(A_s)$ and $G = F - eA_s = \Sigma\, q_i A_i$. Then

$$n(G - L) = n(F - eA_s - L) = n(F - L) - (F - L, A_s)^2/n(A_s)\,,$$

which shows that $n(G - L) < n\,(F - L)$.

To prove the sufficiency, we show that for every set $q_1, \cdots, q_r$ different from $p_1, \cdots, p_r$ then $n(G - L) > n(F - L)$, where $G = \Sigma\, q_i A_i$. Let $s_i = q_i - p_i$, and $H = \Sigma\, s_i A_i$. Then $G = F + H$. Now if $(F - L, A_i) = 0$ for every i, it follows that

$$(F - L, H) = \sum_{i=1}^{r} (F - L, A_i) s_i = 0\,.$$

Thus

$$n(G - L) = n(F - L) + n(H)\,.$$

Since $n(H) > 0$, we have $n(G - L) > n(F - L)$.

The preceding theorem does not require the linear independence of the vectors $A_1, \cdots, A_r$. By Theorem 7a and 7b we see that it is necessary and sufficient for the set $A_1, \cdots, A_r$ to be linearly independent in order to solve the equations $(F - L, A_i) = 0$, $(i = 1, 2, \cdots, r)$, or

$$\begin{aligned} &(A_1, A_1)p_1 + (A_1, A_2)p_2 + \cdots + (A_1, A_r)p_r = (A_1, L) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &(A_r, A_1)p_1 + (A_r, A_2)p_2 + \cdots + (A_r, A_r)p_r = (A_r, L)\,. \end{aligned} \tag{3.9}$$

[6] The proof is based on the same type of reasoning as used by Jackson. See Dunham Jackson's *Theory of Approximation*, pp. 151–152.

If $A_1, \cdots, A_r$ are linearly independent, the conclusion in Theorem 8° can be deduced from Theorem 8. For by Theorem 7a) $A_i = \sum_t s_{it}C_t$, and hence

$$(F - L, A_i) = \left(F - L, \sum_t s_{it}C_t\right) = \sum_t s_{it}(F - L, C_t) = 0 .$$

Also, Theorem 8 can be deduced from Theorem 8°.

IV. Matrices and Their Reciprocals

An ordered array of numbers of the form

$$\alpha = (a_{ij}) = \begin{bmatrix} a_{11}, & a_{12}, & \cdots & a_{1m} \\ a_{21}, & a_{22}, & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \cdots & a_{nm} \end{bmatrix} \tag{4.1}$$

is a matrix. If we write $a(i, j) = a_{ij}$, then the array of numbers (4.1) may be considered as a function of two variables i, j on the ranges of positive integers $(1, 2, \cdots, n)$, $(1, 2, \cdots, m)$.[7] Thus a vector is a special instance of a matrix. We shall use Greek letters to denote matrices throughout this paper unless otherwise specified. When $n = m$, i.e. the number of rows is the same as the number of columns, we have a square matrix. Associated with every n-row square matrix, κ, a determinant can be defined, and for simplicity, we shall adopt the following notation:

$$D(\kappa) = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} .$$

An identity matrix, denoted by $\delta = (d_{ij})$, is a square matrix of which the elements in the principal diagonal are 1 and elsewhere 0, i.e. $d_{ij} = 0$ $(i \neq j)$, $d_{ii} = 1$. A zero matrix, indicated by ω, is one such that every one of its elements is 0. The transposed matrix, α', of α is formed by interchanging the rows and columns. We say two matrices $\alpha = (a_{ij})$ and $\beta = (b_{ij})$ are equal in case $a_{ij} = b_{ij}$ for every i, j. A matrix α is symmetric in case $\alpha' = \alpha$. The i^{th} column of α is indicated by $\alpha(., i)$, the i^{th} row of β by $\beta(i, .)$ and the element in the i^{th} row and j^{th} column by $\alpha(i, j)$. Hence $\alpha(i, j) = a_{ij}$.

ADDITION: Let α be a matrix given by (1) and $\beta = (b_{ij})$ a matrix of the same number of rows and columns as α. Then

$$\alpha + \beta = (a_{ij} + b_{ij}) .$$

[7] E. H. Moore defines a matrix as a function of two variables.

We note that $\alpha + \beta = \beta + \alpha$. If γ is a matrix of the same number of rows and columns as α, then $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

MULTIPLICATION: Let $\alpha = (a_{ij})$ be defined by (1), and $\beta = (b_{jk})$ be a matrix of m row and r columns, then the product $\pi = \alpha\beta$ is defined by

$$\pi = (p_{ik}) = \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right).$$

Thus π is a matrix of m rows and r columns.

The multiplication of two matrices is not necessarily commutative.

If α is a matrix of n rows and m columns, β of m rows and r columns, and γ of r rows and s columns, then $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. If α is a matrix of n rows and m columns, and β, γ are matrices of m rows and r columns, then $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

SCALAR MULTIPLICATION: Let s be a number, and α be a matrix of n rows and m columns, then

$$s \cdot \alpha = (sa_{ij}) = \alpha \cdot s .$$

Let δ_s denote a square matrix of n rows in which the elements in the principal diagonal are s, and 0 elsewhere. Then $\delta_s = s\delta$, where δ is an n row identity matrix. We note from the associative law of multiplication that

$$s\alpha = \delta_s \cdot \alpha = \alpha \cdot \delta_s .$$

In particular, let $s = -1$, then we have -1α. For convenience, we write $-\alpha = -1\alpha$. From the definition of addition, we obtain a definition of subtraction for two matrices of the same number of rows and columns.

RECIPROCALS OF MATRICES: Let α be a matrix of n rows and m columns. Then a matrix α^{-1} of m rows and n columns is said to be a reciprocal of α in case

$$\alpha \cdot \alpha^{-1} = \delta^n , \quad \text{and} \quad \alpha^{-1} \cdot \alpha = \delta^m ,$$

where δ^n, δ^m are identity matrices of order n, m respectively. If a matrix α has a reciprocal α^{-1}, we can prove α^{-1} is unique. It can be shown that *when α has a reciprocal, it must be a square matrix.*[8]

A matrix is said to be non-singular in case it has a reciprocal, otherwise it is said to be singular.[9] It is evident that every zero matrix is singular, and an identity matrix is non-singular.

Suppose α is a square matrix of order n. Let us denote the cofactor of the element a_{ij} of α by e_{ji}. Then

$$\epsilon = (e_{ij}) = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ \cdots & \cdots & \cdots \\ e_{n1} & \cdots & e_{nn} \end{pmatrix}$$

is called the adjoint matrix of α.

---

[8] For the proof of this statement, see Moore, *Vector, Matrices, and Quaternions.*

[9] This definition is due to E. H. Moore.

If α is symmetric, then ϵ is also symmetric. Since $a_{i1}e_{1j} + \cdots + a_{in}e_{nj} = D(\alpha)$ or 0 according as $i = j$ or $i \neq j$, we secure the following:

THEOREM 9. *Let α be a square matrix and ϵ its adjoint, then*

$$\alpha\epsilon = \epsilon\alpha = [D(\alpha)]\delta .$$

THEOREM 10. *If the determinant of α is different from zero, then there exists a reciprocal α^{-1}, and $\alpha^{-1} = adj\ \alpha/D(\alpha)$.*

This theorem follows from theorem 5.

The converse of Theorem 6 is also true.

V. Symmetric Matrices of Positive Type[10]

Let $\alpha = (a_{ij})$ be a matrix of n rows and m columns; and let $\sigma = (k_1, \cdots, k_n)$ and $\rho = (h_1, \cdots, h_m)$ be integers among the sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. The subsets σ and ρ may be equal to the whole sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. Then

$$\alpha(\sigma, \rho) = \begin{Vmatrix} a_{k_1h_1} & \cdots & a_{k_1h_m} \\ \cdots & \cdots & \cdots \\ a_{k_nh_1} & & a_{k_nh_m} \end{Vmatrix} \tag{3}$$

is called a minor of α. In notation we write this minor as $\alpha(\sigma, \rho)$ indicating the ranges to be σ and ρ.

The minor $\alpha(-\sigma, -\rho)$, which is obtained by striking out all the k_i^{th} $(i = 1, \cdots, m)$ columns and h_j^{th} $(j = 1, \cdots, m)$ rows from α, is called the complementary minor of α (σ, ρ).

If α is a square matrix of order n, then $\alpha(\sigma, \sigma)$ is called a principal minor of α.

Let α and β be matrices of n rows and m columns; and let σ, ρ have the same meaning as above. Then $\alpha(\sigma, \rho)$, $\beta(\sigma, \rho)$ are called corresponding minors in α, β respectively.

A symmetric matrix $\alpha = (a_{ij})$ of order n is said to be of *positive type* in case the determinant of every principal minor of α is positive, and is said to be of *properly positive type* in case the determinant of every principal minor of α is greater than zero.

COROLLARY V1. *Every element in the principal diagonal of a positive, symmetric matrix is positive.*

For, let σ consist of a single integer i, then $\alpha(\sigma, \sigma) = a_{ii} \geqq 0$.

COROLLARY V2. *If a symmetric matrix is properly positive, then every element in the principal diagonal is greater than* 0.

THEOREM 11. *If a symmetric matrix α of order n is (properly) positive, then its adjoint matrix ϵ is also symmetric and (properly) positive.*

---

[10] We follow the terminology of E. H. Moore. Moore developed this notion quite extensively.

The symmetry of ϵ is evident. Let σ be a subset of $(1, \cdots, n)$ and let p be the number of integers in σ. Consider any principal minor $\epsilon(\sigma, \sigma)$ in the adjoint matrix ϵ. By a theorem in the theory of determinants, we have[11]

$$D[\epsilon(\sigma, \sigma)] = (-1)^{2k} \cdot D[\alpha(-\sigma, -\sigma)] \cdot [D(\alpha)]^{p-1},$$

where k is an integer depending on the set σ. By hypothesis α is positive (properly positive); hence $D[\alpha(-\sigma, -\sigma)]$ and $[D(\alpha)]^{p-1}$ are positive (greater than 0), and it follows that $D[\epsilon(\sigma, \sigma)]$ is positive (greater than 0).

THEOREM 12. *If a symmetric matrix is properly positive, then $D(\alpha)$ is different from zero, and α has a reciprocal α^{-1}, which is also symmetric and properly positive.*

For take σ to be the whole set $(1, \cdots, n)$ in the definition of proper positiveness, and we see that $D(\alpha) \neq 0$. The theorem now follows from Theorems 10 and 11.

VI. Gramian Matrices

In this section, we shall study the matrices of the normal equations (3.9). The main result is that if the set of observations $A_1, \cdots, A_r$ is linearly independent, then the matrix (called Gramian matrix) is properly positive and has a reciprocal which is also properly positive.

THEOREM 13. *Let $A_1, \cdots, A_r$ be a set of vectors, and let $B_1, \cdots, B_r$ be the orthogonalized set of vectors. Then the matrix*

$$\zeta(A_1, \cdots, A_r) = \begin{pmatrix} (A_1, A_1) & \cdots & (A_1, A_r) \\ \cdots & \cdots & \cdots \\ (A_r, A_1) & \cdots & (A_r, A_r) \end{pmatrix} \tag{6.1}$$

has the following properties:

13.1) *symmetry*

13.2) $D[\zeta(A_1, \cdots, A_r)] = n(B_1)n(B_2) \cdots n(B_r)$,

13.3) *positiveness.*

A matrix of the form (6.1) is called a Gramian matrix.

In fact, the symmetric property follows from the fact that $(A_i, A_j) = (A_j, A_i)$ for every i, j.

We shall prove 13.2) by induction. For $r = 1$, we have by Theorem 5

$$(A_1, A_1) = (B_1, B_1) = n(B_1).$$

Assume the equality is true for $r = t$, we shall show it is true for $r = t + 1$. The $(t + 1)$-row determinant is as follows:

$$D[\zeta(A_1, \cdots, A_r)] = \begin{vmatrix} (A_1, A_1) & \cdots & (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots & \cdots & \cdots & \cdots \\ (A_1, A_t) & \cdots & (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, A_{t+1}) & \cdots & (A_t, A_{t+1}) & (A_{t+1}, A_{t+1}) \end{vmatrix}. \tag{6.2}$$

[11] In case $\sigma = (1, \cdots, n)$, $-\sigma$ is a null class Λ (a class which contains no element); then we define $D[\alpha(-\sigma, -\sigma)] = 1$. For the proof of this theorem, see Bocher, p. 31.

By Theorem 5, there exist constants $s_i (i = 1, \cdots, t)$ such that

$$A_{t+1} = B_{t+1} + \sum_{i=1}^{t} s_i A_i .$$

Substituting this value into the last row, we find the element in the i^{th} column is

$$(A_i, A_{t+1}) = \left(A_i, B_{t+1} + \sum_{j=1}^{t} s_j A_j\right) = (A_i, B_{t+1}) + \sum_{j=1}^{t} s_j (A_i, A_j)$$
$$(i = 1, \cdots, t, t+1).$$

The second term on the right is a linear combination of the first t elements in the i^{th} column of the determinant (6.2) and hence by the theory of determinants,[12] we secure

$$D[\zeta(A_1, \cdots, A_{t+1})] = \begin{vmatrix} (A_1, A_1) & \cdots (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (A_1, A_t) & \cdots (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, B_{t+1}) & \cdots (A_t, B_{t+1}) & (A_{t+1}, B_{t+1}) \end{vmatrix}.$$

By Theorem 5, we find that $(A_i, B_{t+1}) = 0$ for $i = 1, \cdots, t$, and $(A_{t+1}, B_{t+1}) = (B_{t+1}, B_{t+1})$, and hence the preceding determinant reduces to a form in which the first t elements in the $(t+1)^{\text{th}}$ row are zero. Thus

$$D[\zeta(A_1, \cdots, A_{t+1})] = \begin{vmatrix} (A_1, A_1) & \cdots (A_1, A_t) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ (A_1, A_t) & \cdots (A_t, A_t) \end{vmatrix} \cdot n(B_{t+1})$$
$$= n(B_1) n(B_2) \cdots n(B_t) n(B_{t+1})$$

which proves 13.2).

Consider any subset $\sigma = (k_1, \cdots, k_m)$ of the set $(1, \cdots, r)$. By the same argument as above, we find that the determinant of any principal minor

$$\begin{vmatrix} (A_{k_1}, A_{k_2}) & \cdots (A_{k_1}, A_{k_m}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (A_{k_m}, A_{k_1}) & \cdots (A_{k_m}, A_{k_m}) \end{vmatrix} = n(B_{k_1}) \cdots n(B_{k_m}). \tag{6.3}$$

By Theorem 1, the number on the right is positive. Thus the matrix ζ is positive.

THEOREM 14. *The following three assertions are equivalent*:

14.1) *the set* $A_1, \cdots, A_r$ *is linearly independent*;

14.2) *the Gramian matrix* (6.1) *is properly positive*;

14.3) *The determinant of the Gramian matrix* (6.1) *is different from zero.*

We shall prove that 14.1) implies 14.2); 14.2) implies 14.3); and 14.3) implies 14.1). We thus prove the three statements are equivalent.

---

[12] Dickson, *First Course in the Theory of Equations* (1922), p. 113.

Let $B_1, \cdots, B_r$ be the orthogonalized set of the set $A_1, \cdots, A_r$. Since the set $A_1, \cdots, A_r$ is linearly independent, then every subset

$$A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$$

is also linearly independent, and hence $n(B_{k_i}) > 0$ for $i = 1, 2, \cdots, m$. By the same argument as given in the demonstration of Theorem 11, we find that the determinant of any principal minor (6.3) is greater than zero. This proves the matrix (6.1) is properly positive.

If the matrix (6.1) is properly positive, then by Theorem 10 the determinant of (6.1) is different from zero.

To prove 14.3) implies 14.1), suppose $k_i (i = 1, \cdots, r)$ are such that

$$k_1 A_1 + \cdots + k_r A_r = 0.$$

Then

$$(k_1 A_1 + \cdots + k_r A_r, A_i) = k_1(A_1, A_i) + \cdots + k_r(A_r, A_i) = 0$$

for $i = 1, \cdots, r$. Since $(A_i, A_j) = (A_j, A_i)$, and $D(\zeta) \neq 0$, the set of constants k_i must be all equal to 0.[13]

From Theorem 14, and Theorem 10, we may state the following

COROLLARY: *If the set of observations $A_1, \cdots, A_r$ is linearly independent, then the Gramian matrix ζ has a reciprocal which is properly positive.*

VII. Gauss Method of Substitution

LEMMA 7.1) *Let $\varphi = (s_{ij})$ be an r-row symmetric matrix such that $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the following properties:*

a) *$r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i;*

b) *the first minor of r_{11} is symmetric;*

c) *the determinant of every principal minor in ψ of the form*

$$\begin{vmatrix} s_{11} & s_{1k_2} & \cdots & s_{1k_m} \\ 0 & r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & r_{k_2k_m} & \cdots & r_{k_mk_m} \end{vmatrix}, \qquad (2 \leqq k_2 \leqq \cdots \leqq k_m \leqq r) \tag{7.1}$$

is equal to the determinant of the corresponding principal minor in φ.

To prove this lemma, let us define

$$\tau = \delta + F_1 \cdot D_1, \tag{7.2}$$

where D_1 is the first row of an r-row identity matrix δ, and $F_1(1) = 0$,

$$F_1(n) = -s_{1n}/s_{11} \qquad (n > 1).$$

(Thus $F_1 D_1$ is an r-row square matrix in which the first column is F_1 and everywhere else 0.) It is clear that τ thus defined is a square matrix of order r, and

[13] See footnote 5.

$D(\tau) = D(\delta + F_1 D_1) = 1$. By multiplication of these two matrices, $\tau\varphi$, we obtain a new matrix such that $r_{11} = s_{11}$, $r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i, and further

$$r_{ij} = s_{ij} - s_{1i} \cdot s_{1j}/s_{11} \text{ for } i > 1,\ j > 1. \tag{7.3}$$

To prove property (b), we note that $s_{ij} = s_{ji}$, since φ is symmetric. Thus for $i > 1, j > 1$, we note from 7.3) that

$$r_{ij} = s_{ij} - s_{1i}s_{1j}/s_{11} = s_{ji} - s_{1j}s_{1i}/s_{11} = r_{ji}.$$

For the proof of the last property, we note that the corresponding minor of (7.1) in φ is of the form

$$\begin{bmatrix} s_{11} & s_{1k2} & \cdots & s_{1km} \\ s_{1k2} & s_{k2k2} & \cdots & s_{k2km} \\ \cdots & \cdots & \cdots & \cdots \\ s_{1km} & s_{k2km} & \cdots & s_{kmkm} \end{bmatrix}. \tag{7.4}$$

Since φ is symmetric, we have by (7.3),

$$r_{k_i k_j} = s_{k_i k_j} - s_{1k_i}s_{1k_j}/s_{11} \qquad (i > 1,\ j > 1),$$

$$0 = s_{k_i 1} - s_{1k_i}s_{11}/s_{11} \qquad (i > 1).$$

Thus by a theorem in the theory of determinants, the determinants of (7.1) and (7.4) are equal.

LEMMA 7.2) *Let $\varphi = (s_{ij})$ $(i, j = 1, \cdots, r)$ be a symmetric matrix of positive type, and $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the properties stated in Lemma 7.1) and furthermore the minor of r_{11} in 7.1) is of positive type.*

To prove the positiveness of the minor of r_{11}, let the determinant of any one of its principal minors be

$$M_1 = \begin{vmatrix} r_{k2k2} & \cdots & r_{k2km} \\ \cdots & \cdots & \cdots \\ r_{k2km} & \cdots & r_{kmkm} \end{vmatrix} \qquad (2 \leqq k_1 \leqq \cdots \leqq k_m \leqq r),$$

where $r_{k_i k_j} = r_{k_j k_i}$ $(i, j = 2, \cdots, m)$ due to the symmetry. Now consider the bordered determinant

$$M_2 = \begin{vmatrix} r_{11} & r_{1k2} & \cdots & r_{1km} \\ 0 & r_{k2k2} & \cdots & r_{k2km} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & r_{k2km} & \cdots & r_{kmkm} \end{vmatrix}$$

which by property (a) in Lemma 7.1) gives $M_2 = r_{11}M_1 = s_{11}M_1$. By property (c) in Lemma 7.1), M_2 is equal to the determinant of the form (7.4), which by hypothesis is positive. Thus $s_{11}M_1 \geqq 0$. Since $s_{11} > 0$, we conclude that $M_1 = M_2/s_{11} \geqq 0$.

LEMMA 7.3). *Let* $\varphi = (s_{ij})$ $(i, j = 1, 2, \cdots, r)$ *be a symmetric matrix of properly positive type. Then there exists an* r*-row square matrix* τ *whose determinant is unity such that* $\psi = (r_{ij}) = \tau\varphi$ *has the properties stated in Lemma* 7.1) *and furthermore the minor of* r_{11} *in* ψ *is properly positive.*

Since φ is properly positive, we find that $s_{11} > 0$. The proof of this lemma is similar to that of Lemma 7.2).

Suppose that the set of observations $A_1, \cdots, A_r$ is linearly independent. Then by Theorem 14, the Gramian matrix (6.1) is symmetric and properly positive, and hence $(A_1, A_1) > 0$. By Lemma 7.3), the matrix (6.1) can be reduced to the form

$$\begin{bmatrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & \cdots\cdots\cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] \cdots & [A_2A_r\cdot 1] \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & [A_2A_r\cdot 1] & [A_3A_r\cdot 1] \cdots & [A_rA_r\cdot 1] \end{bmatrix} \tag{7.5}$$

where

$$[A_1A_t\cdot 0] = (A_1, A_t) = [A_tA_1\cdot 0] \qquad (t = 1, \cdots, r)$$

$$[A_tA_s\cdot 1] = \frac{[A_1A_1\cdot 0][A_tA_s\cdot 0] - [A_1A_t\cdot 0][A_1A_s\cdot 0]}{[A_1A_1\cdot 0]}.$$

It is evident that $[A_1A_1\cdot 0] = (A_1, A_1) > 0$, since the matrix (6.1) is properly positive. By Lemma 7.3) the value of $D(\zeta)$ and the determinant of (7.5) are equal, and furthermore the minor of the element $[A_1A_1\cdot 0]$ is a symmetric matrix of properly positive type. Thus $[A_2A_2\cdot 1] > 0$, and $[A_tA_s\cdot 1] = [A_sA_t\cdot 1]$.

The minor of $[A_1A_1\cdot 0]$ surely satisfies all the conditions in Lemma 7.3). We may, therefore, apply a transformation of the form (7.2) to the minor of $[A_1A_1\cdot 0]$, and secure another matrix of the same character as (7.5). In other words, we may multiply on the left of the matrix (7.5) by

$$\tau_2 = \delta + F_2D_2 \tag{7.6}$$

where D_2 is the second row of the r row identity matrix δ, and

$$F_2(n) = 0 \qquad (n \leqq 2); \qquad F_2(n) = -\frac{[A_2A_n\cdot 1]}{[A_2A_2\cdot 1]} \qquad (n > 2).$$

In general, let

$$\tau_i = \delta + F_iD_i \qquad (i = 1, \cdots, r-1), \tag{7.7}$$

where D_i is the i^{th} row of the r row identity matrix δ, and

$$F_i(n) = 0 \qquad (n \leqq i); \qquad F_i(n) = -\frac{[A_iA_n\cdot i-1]}{[A_iA_i\cdot i-1]} \qquad (n > i). \tag{7.8}$$

Continuous application of this type of transformation ultimately reduces the matrix (6.1) to the form

$$\eta = \begin{bmatrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & [A_1A_3\cdot 0] & \cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] & \cdots & [A_2A_r\cdot 1] \\ 0 & 0 & [A_3A_3\cdot 2] & \cdots & [A_3A_r\cdot 2] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & [A_rA_r\cdot r-1] \end{bmatrix} \tag{7.9}$$

where

$$[A_tA_s\cdot h] = \frac{[A_hA_h\cdot h-1]\,[A_tA_s\cdot h\text{-}1] - [A_hA_t\cdot h-1]\,[A_hA_s\cdot h\text{-}1]}{[A_hA_h\cdot h-1]} \tag{7.9$_1$}$$

$$(t, s = 1, \cdots, r; \qquad 0 \leqq h \leqq sm(t, s)).^{14}$$

In the matrix (7.9), we see by virtue of Lemma 7.3) that $[A_iA_i\cdot i-1] > 0$ for every i, and $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ for every s, t and $0 \leqq h \leqq sm(t, s)$. If $h = sm(t, s)$, then $[A_tA_s\cdot h] = 0$.

Let $\tau = \tau_{r-1}\cdot\tau_{r-2}\cdots\tau_1$. Then by the associative law of multiplication of matrices, we see that

$$\eta = (\tau_{r-1}\cdots\tau_1)\zeta = \tau\zeta. \tag{7.10}$$

Thus we prove

THEOREM 15. *If the set of vectors $A_1, \cdots, A_r$ is linearly independent, then there exists a square matrix τ of order such that $\tau\zeta$ is of the form* (7.9) *where all elements below the principal diagonal are* 0; *every element in the principal diagonal $[A_iA_i\cdot i-1]$ $(i = 1, \cdots, r)$, is greater than zero; and $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ for $s, t = 1, \cdots, r$, and $h < sm(t, s)$. Furthermore the determinants of the matrices* (6.1) *and* (7.9) *are equal.*

We now prove the following lemma which will be useful in the later section.

LEMMA 7.4). *If $[A_iA_i\cdot i-1]$ is different from zero for every $i \geqq 0$, then for every pair of integers (s, t), where $s, t = 1, \cdots, r$, and $n \leqq sm(t, s)$, we have*

a) $$[A_tA_s\cdot n] = (A_t, A_s) - \sum_{i=1}^{n-1} \frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}\,[A_iA_s\cdot i-1].$$

b) $$[A_t(A_s + A_u)\cdot n] = [A_tA_s\cdot n] + [A_tA_u\cdot n], \quad (u = 1, \cdots, r).$$

c) $$[(cA_t)A_s\cdot n] = c\,[A_tA_s\cdot n], \qquad (c = a \text{ constant}).$$

To prove a), take every pair (s, t). We find the lemma is true for $n = 0$. Assuming it is true for every (s, t) and for $n = h < sm(s, t)$, we find that $h + 1 \leqq sm(s, t)$, and

$$(A_t, A_s) - \sum_{i=1}^{h+1} \frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}\,[A_iA_s\cdot i-1]$$

[14] sm (s, t) read "the smaller one of (t, s)."

$$= (A_t, A_s) - \sum_{i=1}^{h} \frac{[A_iA_t \cdot i-1]}{[A_iA_i \cdot i-1]} [A_iA_s \cdot i-1] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h]$$

$$= [A_tA_s \cdot h] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h] = [A_tA_s \cdot h+1],$$

for every s, t.

Parts b) and c) are true for $n = 0$. Now make use of the equality in a) and prove by induction.

VIII. Gauss's Method of Substitution and its Relation to Gramian Schmidt's Orthogonalization Process

Let us write the set of observations in the form:

$$\alpha \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{pmatrix}.$$

Let the orthogonalized set also be written in the form

$$\beta = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{r1} & \cdots & b_{rn} \end{pmatrix}.$$

From Theorems 5 and 6, we find that there exists a transformation κ given by an r-row square matrix such that $\beta = \kappa\alpha$. Thus by the associative law of multiplication of matrices, we have

$$\beta\alpha' = (\kappa\alpha)\alpha' = \kappa(\alpha\alpha').$$

Now the matrix $\alpha\alpha'$ is the Gramian matrix (6.1). Thus

$$\beta\alpha' = \kappa\zeta. \tag{8.1}$$

The composite matrix $\beta\alpha'$ is of the form

$$\begin{bmatrix} (B_1, A_1)(B_1, A_2) \cdots (B_1, A_r) \\ (B_2, A_1)(B_2, A_2) \cdots (B_2, A_r) \\ \cdots \cdots \cdots \cdots \\ (B_r, A_1)(B_r, A_2) \cdots (B_r, A_r) \end{bmatrix}. \tag{8.2}$$

By Theorems 5 and 6, we note that $(B_s, A_t) = 0$ for $s > t$, and $(B_s, A_s) = (B_s, B_s)$ for every s. Thus the preceding matrix can be written in the form

$$\begin{bmatrix} (B_1, B_1)(B_1, A_2)(B_1, A_3) \cdots (B_1, A_r) \\ 0 \quad (B_2, B_2)(B_2, A_3) \cdots (B_2, A_r) \\ \cdots \cdots \cdots \cdots \\ 0 \quad 0 \quad 0 \quad \cdots (B_r, B_r) \end{bmatrix}. \tag{8.3}$$

We have proved the following theorem:

THEOREM 16. *Let $A_1, \cdots, A_r$ be a set of vectors, and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then there exists a square r-row matrix κ such that $\beta = \kappa\alpha$, and $\kappa\alpha\alpha'$ is a matrix of the form* (8.3) *where all the elements below the principal diagonal are zeros and every element in the principal diagonal is positive. If the set $A_1, \cdots, A_r$ is linearly independent, then every element in the principal diagonal is greater than zero.*

THEOREM 17. *Let $A_1, \cdots, A_r$ be a set of vectors and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then $D(\beta\alpha') = D(\alpha\alpha')$.*

For by equations (2.1), we note that $D(\kappa) = 1$. Thus

$$D(\beta\alpha') = D(\kappa\alpha\alpha') = D(\kappa)D(\alpha\alpha') = D(\alpha\alpha') .$$

THEOREM 18. *If the set of vectors, $A_1, \cdots, A_r$ is linearly independent, the matrix κ arising from Gram-Schmidt's orthogonalization process is identical with the matrix τ defined by* (7.10).

To prove this theorem, we first establish the following

LEMMA 8.5): *If the set $A_1, \cdots, A_r$ be linearly independent, and $B_1, \cdots, B_r$ be the orthogonalized set, then for every t, h, we have*

$$(B_h, A_t) = [A_h A_t \cdot h - 1] .$$

By Theorem 10, the set B_i is linearly independent, and hence $n(B_i) > 0$ for every i. The lemma is evidently true for every t and $h = 1$. Assuming it is true for every t and $h = s$, we shall prove it is also true for every t and $h = s - 1$. Now

$$B_{s+1} = A_{s+1} - \sum_{i=1}^{s} \frac{(A_s, B_i)}{(B_i, B_i)} B_i = A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i .$$

Thus by the linear property of (,) we secure, for every t

$$\begin{aligned}
(B_{s+1}, A_t) &= \left(A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i, A_t \right) \\
&= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} (B_i, A_t) \\
&= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} [A_i A_t \cdot i - 1] \\
&= [A_{s+1} A_t \cdot s]
\end{aligned}$$

by virtue of lemma 4.4).

From this lemma, we conclude at once that the matrices (7.9) and (8.3) are equal. Thus by (8.1), we have

$$\kappa\zeta = \beta\alpha' = \tau\zeta, \quad \text{or} \quad (\kappa - \tau)\zeta = \omega .$$

Since ζ is non-singular (by Theorem 12), we have

$$\omega = (\kappa - \tau)\zeta\zeta^{-1} = (\kappa - \tau)\delta = \kappa - \tau,$$

which proves the theorem.

From Lemma 8.5), we have

LEMMA 8.6). Let $L = (l_1, \cdots, l_n)$. Suppose the set $A_1, \cdots, A_r$ to be linearly independent, and $B_1, \cdots, B_r$ to be the orthogonalized set. Then for every h

$$[A_h L \cdot h - 1] = (B_h, L) .$$

Theorems 16, 17, and 18 furnish us a new method for finding the most probable values of the unknowns in the theory of least squares. The formulation of the system of normal equations may be omitted in this new procedure, which may be described briefly as follows: After we obtain a set of observations $A_1, \cdots, A_r$, we orthogonalize this set by means of Gram-Schmidt's process. Let L be a non-zero vector. The product

$$\begin{pmatrix} b_{11} \cdots b_{1n} \\ \cdots\cdots\cdots \\ b_{r1} \cdots b_{rn} \end{pmatrix} \quad \begin{pmatrix} a_{11} \cdots a_{r1}, & -l_1 \\ \cdots\cdots\cdots\cdots & \\ a_{1n} \cdots a_{rn}, & -l_n \end{pmatrix}$$

will give us the result as desired by Gauss's method of substitution.

ACADEMIA SINICA,
PEIPING, CHINA.

A NOTE ON THE ANALYSIS OF VARIANCE[1]

By Solomon Kullback

By considering a set of independent items classified in some relevant manner into N sets of s items each, and by the use of a dispersion theorem of Prof. J. L. Coolidge,[2] Prof. H. L. Rietz[3] arrives at estimates of variance, used by Dr. R. A. Fisher, without making use of arguments involving the number of degrees of freedom of the items concerned.

By proceeding along the lines followed by Coolidge and Rietz but considering a set of independent items classified into N sets of $s_i (i = 1, 2, \cdots, N)$ items each, we shall arrive at certain other important results of R. A. Fisher[4] in his analysis of variance.

The theorem referred to above is as follows: If n independent quantities $y_1, y_2, \cdots, y_n$ be given, their expected values being $a_1, a_2, \cdots, a_n$, while the expected values of their squares are $A_1, A_2, \cdots, A_n$, respectively, and if we agree to set $y = (1/n) \sum_{i=1}^{n} y_i$, $a = (1/n) \sum_{i=1}^{n} a_i$, then the expected value of the variance, $(1/n) \sum_{i=1}^{n} (y_i - y)^2$ is

$$\frac{1}{n}\left[\frac{n-1}{n}\sum_{i=1}^{n}(A_i - a_i^2) + \sum_{i=1}^{n}(a_i - a)^2\right]. \tag{1}$$

Suppose a set of independent items has been classified in some relevant manner into N sets of s_i $(i = 1, 2, \cdots, N)$ items each as follows:

$$\begin{array}{lllll} x_{11}, & x_{12}, & \cdots, & x_{1s_1}, & \bar{x}_1 \\ x_{21}, & x_{22}, & \cdots, & x_{2s_2}, & \bar{x}_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{N1}, & x_{N2}, & \cdots, & x_{Ns_N}, & \bar{x}_N \\ & & & & \bar{x} \end{array} \tag{2}$$

where $\bar{x}_i (i = 1, 2, \cdots, N)$ is the arithmetic mean of the i^{th} set and $\bar{x}$ the mean of the pooled sample of $s = s_1 + s_2 + \cdots + s_N$ items.

We shall assume that the set (2) is statistically homogeneous in the sense that,

---

[1] Presented to the American Mathematical Society, February 23, 1935.

[2] Bulletin Am. Math. Soc., Vol. 27 (1921) p. 439.

[3] Bulletin Am. Math. Soc., Vol. 38 (1932) pp. 731–735.

[4] Proceedings of the International Math. Congress, Toronto, 1924, Vol. 2, p. 802 ff.

using $E(\ )$ for the expected value of the expression in the parenthesis, we may let $E(x_{ij}) = a$, $E(x_{ij}^2) = A$, $(i = 1, 2, \cdots, N, j = 1, 2, \cdots, s_i)$. Then, using (1)

$$E\left(\sum_{j=1}^{s_i} (x_{ij} - \bar{x}_i)^2\right) = (s_i - 1)(A - a^2). \tag{3}$$

Summing (3) from $i = 1$ to N, we have

$$E\left(\sum_{i=1,\ j=1}^{N,\ s_i} (x_{ij} - \bar{x}_i)^2\right) = (A - a^2) \sum_{i=1}^{N} (s_i - 1) = (s - N)(A - a^2). \tag{4}$$

Similarly, by using (1)

$$E\left(\sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2\right) = \frac{N-1}{N} \sum_{i=1}^{N} s_i [E(\bar{x}_i^2) - a^2]. \tag{5}$$

But[5]

$$E(\bar{x}_i^2) - a^2 = E(\bar{x}_i - a)^2, \qquad \text{and} \tag{6}$$

$$E(\bar{x}_i - a)^2 = (A - a^2)/s_i, \qquad \text{therefore} \tag{7}$$

$$E\left(\sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2\right) = (N - 1)(A - a^2). \tag{8}$$

Similarly by using (1)

$$E\left(\sum_{i=1,\ j=1}^{N,\ s_i} (x_{ij} - \bar{x})^2\right) = (s - 1)(A - a^2). \tag{9}$$

Thus, in a statistically homogeneous set of items, classified as in (2), the following estimates of Variance have the same expected value:

$$V = \frac{S}{s-1}, \qquad \text{where} \qquad S = \sum_{i=1,\ j=1}^{N,\ s_i} (x_{ij} - \bar{x})^2$$

$$V_i = \frac{S_i}{s-N}, \qquad \text{where} \qquad S_i = \sum_{i=1,\ j=1}^{N,\ s_i} (x_{ij} - \bar{x}_i)^2 \tag{10}$$

$$V_{\bar{x}} = \frac{S_{\bar{x}}}{N-1}, \qquad \text{where} \qquad S_{\bar{x}} = \sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2.$$

These estimates are used in applying the analysis of variance to the study of the correlation ratio, η, for uncorrelated material, where $\eta^2 = S_{\bar{x}}/S$.

OFFICE OF THE CHIEF SIGNAL OFFICER,
WASHINGTON, D. C.

[5] Rietz, H. L., loc. cit. p. 733.

A PROBLEM INVOLVING THE LEXIS THEORY OF DISPERSION

By Walter A. Hendricks

The attention of the author was recently directed to a study of the hatchability of chicken eggs at the U. S. Animal Husbandry Experiment Station, Beltsville, Maryland. It was necessary to find the average hatchability of the fertile eggs incubated for each of a number of lots of birds and the corresponding standard errors of those averages.

It was very apparent that some methods for computing such values, in common use at the present time, do not give satisfactory results. This is due to the fact that the fertile eggs produced by different birds vary considerably with respect to hatchability as well as with respect to number of eggs available for incubation. It seems reasonable to suppose that the variability in hatchability of a number of fertile eggs, produced by a given number of birds, should obey the Lexis law of dispersion. This supposition is based on two hypotheses:

(a) The probability that a fertile egg will hatch is constant for all fertile eggs produced by the same bird during the time interval under consideration.

(b) The probability that a fertile egg will hatch varies from bird to bird.

The reader familiar with the principles of genetics may question the validity of the first of these hypotheses. The probability that a fertile egg will hatch is largely governed by the genes carried by the chromosomes of the ovum of the hen and the sperm of the male bird which fertilized that ovum. The kinds of genes carried by various ova and spermatozoa are not necessarily the same, even when those ova and spermatozoa are produced by the same female and male birds, respectively. However, if we have a sample of a number of fertile eggs produced by the same hen, we are justified in assuming that the proportion of those eggs which will hatch is constant, except for sampling fluctuations, when successive samples of fertile eggs produced by the given hen are incubated, provided, of course, that the eggs in the successive samples were fertilized by the same male bird or birds. The limit approached by the proportion of fertile eggs which hatch as the number of fertile eggs produced by the given hen becomes infinitely large may be defined as the probability that a fertile egg produced by that hen will hatch. It will be recognized that this definition is based on purely academic considerations, since there are physical limitations to the number of fertile eggs which a hen can produce in a given period of time. Hypotheses (a) and (b) are to be interpreted in the light of this definition of the probability that a fertile egg produced by a given bird will hatch.

Let $s_1, s_2, \cdots s_n$ represent the numbers of fertile eggs produced by n birds during a period of time and let $f_1, f_2, \cdots f_n$, respectively, represent the numbers

of chicks obtained from those eggs when the eggs are incubated. Let $p_k = \frac{f_k}{s_k}$ represent the hatchability of the fertile eggs produced by the k^{th} bird.

The squared standard error of p_k is given by the Lexis formula:[1]

$$\sigma^2_{p_k} = \frac{PQ}{s_k} + \frac{s_k - 1}{ns_k} \sum_{t=1}^{n} (P_t - P)^2 \tag{1}$$

in which the P_t represent the respective probabilities that the fertile eggs produced by the n birds will hatch, P is the arithmetic mean of the P_t, and Q is equal to $1 - P$.

The values of the probabilities, P_t, are not known. However, as a first approximation to equation (1) we may write:

$$\sigma^2_{p_k} = \frac{pq}{s_k} + \frac{s_k - 1}{ns_k} \sum_{t=1}^{n} (p_t - p)^2 \tag{2}$$

in which p is the arithmetic mean of the p_t and q is equal to $1 - p$.

The product, pq, can be accepted as a reasonably close approximation to the product, PQ, but the expression, $\sum_{t=1}^{n} (p_t - p)^2$, will, in general, be greater than the expression, $\sum_{t=1}^{n} (P_t - P)^2$. The reason for this is apparent when we consider that if each of these two expressions is divided by n, the former yields an estimate of the squared standard deviation of the p_t while the latter yields an estimate of the squared standard deviation of the P_t. The standard deviation of the p_t will, in general, be greater than that of the P_t because the p_t are more or less imperfect estimates of the P_t and are, therefore, subject to sampling errors from which the P_t are free.

We may write:

$$\frac{1}{n} \sum_{t=1}^{n} (P_t - P)^2 = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 - \sigma_c^2 \tag{3}$$

in which σ_c^2 is an appropriate correction as yet undefined.

Since the p_t would approach the P_t as statistical limits if each of the s_t were made extremely large, it follows that σ_c^2 must approach zero as each of the s_t approaches infinity. Furthermore, if $P_1 = P_2 = \cdots P_n = P$, we must have:

$$\frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 - \sigma_c^2 = 0 \qquad \text{or}$$

$$\sigma_c^2 = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2. \tag{4}$$

[1] The formula as given in this paper is a modification of that given by Rietz, H. L. (1927) in his book, Mathematical Statistics, Open Court Publishing Co., Chicago, which was necessary in order to make it applicable to relative frequencies.

These conditions suggest that σ_c^2 be defined by the equation:

$$\sigma_c^2 = \frac{pq}{n} \sum_{t=1}^{n} \frac{1}{s_t}. \tag{5}$$

If σ_c^2 is so defined, it will obviously approach zero as each of the s_t approaches infinity. Furthermore, it has been shown by Yule[2] that if we have a series of n relative frequencies, such as the p_t under discussion, based on n samples of unequal size, and the probabilities of the occurrence and non-occurrence, respectively, of the particular event under consideration are constant from sample to sample, the squared standard deviation of those relative frequencies is given by a relation such as that used to define σ_c^2 in equation (5). Therefore, the second condition is also satisfied. σ_c^2 may be interpreted as representing that part of the squared standard deviation of the p_t which is due to the unreliability of the p_t as estimates of the P_t.

Therefore, it seems reasonable to write:

$$\frac{1}{n} \sum_{t=1}^{n} (P_t - P)^2 = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 - \frac{pq}{n} \sum_{t=1}^{n} \frac{1}{s_t}. \tag{6}$$

Combining equations (1) and (6), we obtain the following formula for calculating the squared standard error of p_k:

$$\sigma_{p_k}^2 = \frac{pq}{s_k} + \frac{s_k - 1}{n s_k} \left[\sum_{t=1}^{n} (p_t - p)^2 - pq \sum_{t=1}^{n} \frac{1}{s_t} \right]. \tag{7}$$

Since the weight of a measurement is inversely proportional to the square of the standard error of the measurement, we are now in a position to calculate a weighted mean, $\bar{p}$, of the p_t.

$$\bar{p} = \frac{\sum_{t=1}^{n} w_t p_t}{\sum_{t=1}^{n} w_t} \tag{8}$$

in which:

$$w_t = \frac{1}{\sigma_{p_t}^2}. \tag{9}$$

The squared standard error of $\bar{p}$ is given by the familiar formula:

$$\sigma_{\bar{p}}^2 = \frac{\sum_{t=1}^{n} w_t (p_t - \bar{p})^2}{(n - 1) \sum_{t=1}^{n} w_t}. \tag{10}$$

---

[2] Yule, G. Udny, 1927. Introduction to the Theory of Statistics, Charles Griffin and Co., London.

It would seem that $\bar{p}$ may be accepted as a good estimate of the average hatchability of the fertile eggs produced by the given lot of birds, and that equation (10) may be used to obtain a valid estimate of the reliability of $\bar{p}$.

However, the problem is not quite so simple. In the first place, there is usually a small amount of positive correlation between the number of fertile eggs produced by a bird and the hatchability of those eggs. Secondly, as pointed out earlier in this paper, the hatchability of fertile eggs is influenced to some extent by the male birds used to fertilize the eggs. The error involved in neglecting the correlation between hatchability and number of fertile eggs incubated does not seem to be of much importance in those practical problems which have come to the author's attention. The effects of differences among the male birds may be largely obviated in experimental work by frequently transferring male birds from lot to lot during the experimental period.

The best test of the suitability of a particular formula for calculating the standard error of an average is to compare the value of the standard error calculated by means of the formula with the corresponding value obtained by direct calculation from the distribution of a number of such averages obtained under essentially the same conditions. The accompanying table gives the standard error of the weighted average hatchability of fertile eggs calculated for each of four lots of birds by means of equation (10), together with the corresponding values obtained from the distribution of averages. The former are designated as the "predicted" values and the latter are designated as the "observed" values. In the calculation of the "observed" values, the various averages were assigned the same weights which were used in the calculation of the "predicted" values.

Comparison of "predicted" and "observed" standard errors of the weighted average hatchability of fertile eggs, calculated for each of four lots of birds

Lot	$\bar{p}$	Standard error of $\bar{p}$	
		"Predicted"	"Observed"
1	0.7684	0.0287	0.0327
2	0.7115	0.0533	0.0561
3	0.6834	0.0355	0.0379
4	0.7260	0.0615	0.0674

The data used in these calculations involved a total of 74 birds, approximately equally divided among the four lots, and a total of 2,901 fertile eggs which were produced and incubated during an experimental period of 48 weeks. The agreement between the "predicted" and "observed" standard errors of the weighted average hatchability for each lot of birds is excellent. However, the author's experience with biological data tends to make him doubt that such

close agreement will always be found when such data are subjected to the above treatment. The agreement in the present illustration could be less close without indicating that the method of calculating the "predicted" standard errors is unsound.

BUREAU OF ANIMAL INDUSTRY,
U. S. DEPARTMENT OF AGRICULTURE,
WASHINGTON, D. C.

A METHOD FOR DETERMINING THE COEFFICIENTS OF A CHARACTERISTIC EQUATION

PAUL HORST

For the characteristic equation

$$\begin{vmatrix} a_{11} - x & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} - x \end{vmatrix} \equiv (-1)^n(x^n - c_1x^{n-1} + c_2x^{n-2} - \cdots + c_n) \equiv (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \tag{1}$$

it is well known that

$$c_i = A_i$$

where A_i is the sum of all i^{th} order co-axial minors of the determinant

$$A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}. \tag{2}$$

If n exceeds 3 or 4, the process of calculating all possible principal minors is very cumbersome.

But another more systematic method of calculating the c's may be adopted. Suppose we define

$$A^p = \begin{vmatrix} a_{11}^{(p)} & \cdots & a_{1n}^{(p)} \\ \cdots & \cdots & \cdots \\ a_{n1}^{(p)} & \cdots & a_{nn}^{(p)} \end{vmatrix} \tag{3}$$

and

$$\sum_1^n \alpha_i^{(p)} = S_p. \tag{4}$$

It may be proved[1] that

$$S_p = \sum_1^n a_{ii}^{(p)}. \tag{5}$$

But from Newton's identities[2] we have

$$S_p + c_1S_{p-1} + c_2S_{p-2} + \cdots + c_{p-1}S_1 + pc_p = 0. \tag{6}$$

[1] Muir, L. & Metzler, W. H., "A Treatise on the Theory of Determinants," p. 606, ¶ 650 and 651.

[2] Dickson, L. E., "First Course in the Theory of Equations," p. 134, ¶ 106.

Newton's identities are ordinarily employed for calculating the sums of the powers of the roots of a polynomial when the coefficients are known. They may be employed equally well, however, for calculating the coefficients when the sums of the powers are given. Thus by means of equations (5) and (6) the coefficients of (1) may be readily calculated.

If in (2) $a_{ij} = a_{ji}$, the calculation of the successive A^p values is straightforward. The determinant A is used as a constant multiplier so that

$$A \cdot A = A^2, \qquad A \cdot A^2 = A^3, \cdots A \cdot A^{n-1} = A^n$$

and the multiplication is column by column. That is,

$$a_{ij}^{(1+p)} = \sum_{k=1}^{n} a_{ki} a_{kj}^{(p)} .$$

THE GENERALIZED PROBLEM OF CORRECT MATCHINGS

By Dwight W. Chapman

A method common to many experimental and testing procedures in psychology and education is to require an individual to match, as best he can, members of one series of items with members of a second series of quite different items certain of which are in some sense true apposites of items in the first series. Thus the experimental psychology of personality has often investigated the ability of graphologists or laymen to pair samples of handwriting produced by a group of persons with, say, character-sketches of these same persons; and the excess of correct matchings thus produced over the number to be expected by chance has been used as evidence that the expressive movement of handwriting affords characteristics diagnostic of personal traits. Fortunately, the excesses experimentally obtained have often been so large as obviously to exclude the operation of chance alone. But many empirical results show small excesses only; and the interpretation of such findings has not hitherto been subjected to rigid statistical analysis.

The particular statistical problem resident in this experimental procedure is twofold, involving the estimation of the significance of (a) a given number of correct matchings produced by one individual, and (b) a given mean number of correct matchings produced by a group of individuals working with the same material independently.

Furthermore, two cases arise in practice: (1) the two series of items are of equal length, and each item in either series has a true apposite in the other series; or (2) the two series may be of unequal length, in which case the longer series contains not only a true apposite for each item of the shorter series, but, in addition, a certain number of extra, irrelevant items which cannot be correctly matched with any items in the shorter series. I have already given the solution to problems (a) and (b) for case (1).[1] But case (1) forms only a corollary of the more general case (2), to the solution of which this present paper is devoted.

(a) The Significance of a Given Number of Correct Matchings Resulting from a Single Trial

Let there be given a series of u x-items,

$$x_1, x_2, \cdots x_t, \cdots x_u$$

and a series of t y-items,

$$y_1, y_2, \cdots y_t .$$

[1] The Statistics of the Method of Correct Matchings, *Amer. Jour. Psychol.*, 46, 1934, 287-298.

Let $t \leqq u$, and let the first t x-items be in some sense true apposites of the correspondingly numbered y-items, so that if y_j be paired with x_j $(j = 1, 2, \cdots t)$, this pairing will constitute a correct matching.

The first problem which arises is that of determining the probability that a single random arrangement of the t y-items against t of the x-items will result in exactly s $(= 0, 1, 2, \cdots t)$ correct matchings.

We begin by putting the first s y-items in correspondence with their apposite x-items. Then the number of arrangements of the t y-items in which only these s are correctly matched is the number of arrangements of the remaining $t - s$ y-items against the remaining $u - s$ x-items such that no correct matchings occur. With respect to these items, let

n = the number of all possible arrangements,

$n(Y_j)$ = the number of arrangements such that at least the j^{th} item is correctly matched with its apposite,

$n(Y_jY_k)$ = the number of arrangements such that at least both the j^{th} and k^{th} items are matched with their apposites, etc.;

and let

$n(\bar{Y}_j)$ = the number of arrangements such that at least the j^{th} item is not matched with its apposite,

$n(\bar{Y}_j\bar{Y}_k)$ = the number of arrangements such that at least the j^{th} and k^{th} items are not matched with their apposites, etc.

We have then to evaluate the expression $n(\bar{Y}_{s+1}\bar{Y}_{s+2} \cdots \bar{Y}_t)$, the number of arrangements of the items remaining, after setting s of them correctly matched, such that no further correct matchings occur.

Now it can be shown that[2]

$$\begin{aligned} n(\bar{Y}_{s+1}\bar{Y}_{s+2} \cdots \bar{Y}_t) = n \\ &- [n(Y_{s+1}) + n(Y_{s+2}) + \cdots + n(Y_t)] \\ &+ [n(Y_{s+1}Y_{s+2}) + n(Y_{s+1}Y_{s+3}) + \cdots + n(Y_{t-1}Y_t)] \\ &- [n(Y_{s+1}Y_{s+2}Y_{s+3}) + \cdots + n(Y_{t-2}Y_{t-1}Y_t)] \\ &+ \cdots \\ &+ (-1)^t\, n(Y_{s+1}Y_{s+2} \cdots Y_t)\,. \end{aligned}$$

The value of the expressions on the right side of this equation can be determined as follows:

---

[2] H. Whitney, A Logical Expansion in Mathematics, *Bull. Amer. Math. Soc.*, 1932, 572–579.

The value of n is the number of ways in which $t - s$ items can be arranged against

$u - s$ items, which is $\dfrac{(u - s)!}{[(u - s) - (t - s)]!} = \dfrac{(u - s)!}{(u - t)!}$.

The value of the first bracket—the number of arrangements of these items such that some one of them is correctly matched—is derived by holding one of the items matched, which can be chosen in $t - s$ ways. This leaves $t - s - 1$ y-items, which can be arranged against the remaining $u - s - 1$ x-items in $(u - s - 1)!/(u - t)!$ ways. The product of these two expressions gives us for the value of the first bracket

$$[n(Y_{s+1}) + \cdots + n(Y_t)] = \frac{(t - s)!(u - s - 1)!}{(u - t)!}.$$

To evaluate the second bracket, we hold two of the $t - s$ items matched, which can be chosen in $(t - s)!/[2!(t - s - 2)!]$ ways. There remains $t - s - 2$ y-items which can be arranged against the remaining $u - s - 2$ x-items in $(u-s-2)!/(u-t)!$ ways. The product of these two expressions gives us

$$[n(Y_{s+1}Y_{s+2}) + \cdots + n(Y_{t-1}Y_t)] = \frac{(t - s)!(u - s - 2)!}{2!(t - s - 2)!(u - t)!}.$$

Continuing thus, we develop the following series for the number of arrangements of t items against u items such that the first s are correctly matched:

$$n(\bar{Y}_{s+1}\bar{Y}_{s+2}\cdots\bar{Y}_t) = \frac{(u - s)!}{(u - t)!} - \frac{(t - s)!(u - s - 1)!}{(u - t)!} + \frac{(t - s)!(u - s - 2)!}{2!(t - s - 2)!(u - t)!}$$
$$- \cdots + (-1)^{t-s}\frac{(t - s)!(u - t)!}{(t - s)!(u - t)!}.$$

In order to express the number of arrangements, $N_{(s)}$, such that *any* s correct matchings occur, we must multiply the above series by $t!/[s!(t - s)!]$, which is the number of ways in which s items can be chosen from t items:

$$N_{(s)} = \frac{t!}{s!(t - s)!}\left[\frac{(u - s)!}{(u - t)!} - \frac{(t - s)!(u - s - 1)!}{(u - t)!}\right.$$
$$\left. + \cdots + (-1)^{t-s}\frac{(t - s)!(u - t)!}{(t - s)!(u - t)!}\right].$$

And in order to obtain the probability that a single random arrangement will result in exactly s correct matchings, we must further divide by $u!/(u - t!)$, which is the total number of ways in which t items can be arranged against u items. Calling this probability $P_{(s)}$, we have then

$$P_{(s)} = \frac{t!\,(u - t)!}{u!s!\,(t - s)!}\left[\frac{(u - s)!}{(u - t)!} - \frac{(t - s)!\,(u - s - 1)!}{(u - t)!}\right.$$
$$\left. + \cdots + (-1)^{t-s}\frac{(t - s)!\,(u - t)!}{(t - s)!\,(u - t)!}\right].$$

Finally, factoring $(t-s)!/(u-t)!$ out of all terms in the bracket, the series simplifies to[3]

$$P_{(s)} = \frac{t!}{s!u!}\left[\frac{(u-s)!}{0!(t-s)!} - \frac{(u-s-1)!}{1!(t-s-1)!} + \frac{(u-s-2)!}{2!(t-s-2)!} - \cdots + (-1)^{t-s}\frac{(u-t)!}{(t-s)!\,0!}\right]. \tag{1}$$

In any practical situation, the significant question is not the probability that exactly s correct matchings shall occur, but the probability of s *or more* correct matchings. Obviously

$$P_{(s \text{ or more})} = P_{(s)} + P_{(s+1)} + \cdots + P_{(t)}.$$

whence, by equation (1),

$$\begin{aligned} P_{(s \text{ or more})} = {} & \frac{t!}{s!u!}\left[\frac{(u-s)!}{0!(t-s)!} - \frac{(u-s-1)!}{1!(t-s-1)!} + \frac{(u-s-2)!}{2!(t-s-2)!} - \cdots + (-1)^{t-s}\frac{(u-t)!}{(t-s)!0!}\right] \\ & + \frac{t!}{(s+1)!u!}\left[\frac{(u-s-1)!}{0!(t-s-1)!} - \frac{(u-s-2)!}{1!(t-s-2)!} + \cdots + (-1)^{t-s-1}\frac{(u-t)!}{(t-s-1)!0!}\right] \\ & + \frac{t!}{(s+2)!u!}\left[\frac{(u-s-2)!}{0!(t-s-2)!} - \cdots + (-1)^{t-s-2}\frac{(u-t)!}{(t-s-2)!0!}\right] \\ & + \cdots \\ & + \frac{t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]. \end{aligned} \tag{2}$$

Or, collecting terms in a form better suited to practical computation from tables of factorials and reciprocals,

$$\begin{aligned} P_{(s \text{ or more})} = \frac{t!}{u!}\Bigg\{ & \frac{(u-s)!}{(t-s)!}\left[\frac{1}{0!s!}\right] \\ & + \frac{(u-s-1)!}{(t-s-1)!}\left[\frac{1}{0!(s+1)!} - \frac{1}{1!s!}\right] \\ & + \frac{(u-s-2)!}{(t-s-2)!}\left[\frac{1}{0!(s+2)!} - \frac{1}{1!(s+1)!} + \frac{1}{2!s!}\right] \end{aligned}$$

[3] In the special case in which the series of x-items and the series of y-items are of the same length, whence $t = u$, equation (1) reduces to

$$P_{(s)} = \frac{1}{s!}\left[\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^{t-s}\frac{1}{(t-s)!}\right].$$

$$+ \cdots$$

$$+ \frac{(u-t)!}{0!}\left[\frac{1}{0!t!} - \frac{1}{1!(t-1)!} + \frac{1}{2!(t-2)!} - \cdots + (-1)^{t-s}\frac{1}{(t-s)!s!}\right]\Bigg\}. \qquad (2')$$

(b) The Significance of a Given Mean Number of Correct Matchings Resulting from n Independent Trials

A frequent practical situation is that in which interest centers on the significance of the mean number of correct matchings achieved by a group of n individuals working independently with the same two series.

In order to determine the probability that the mean number of correct matchings, $\bar{s}$, resulting from n independent trials shall equal or exceed a given value, we are required to describe the distribution of the means of samples of size n drawn at random from a parent population in which the variable is $s(= 0, 1, 2, \cdots t)$ with relative frequencies $P_{(0)}, P_{(1)}, P_{(2)}, \cdots P_{(t)}$, given by equation (1). The tabulation of this parent distribution follows:

Table I: Distribution of s

s	*Relative frequency* $(= P_{(s)})$
0	$\frac{t!}{0!u!}\left[\frac{u!}{0!t!} - \frac{(u-1)!}{1!(t-1)!} + \frac{(u-2)!}{2!(t-2)!} - \frac{(u-3)!}{3!(t-3)!} + \cdots + (-1)^t\frac{(u-t)!}{t!0!}\right]$
1	$\frac{t!}{1!u!}\left[\frac{(u-1)!}{0!(t-1)!} - \frac{(u-2)!}{1!(t-2)!} + \frac{(u-3)!}{2!(t-3)!} - \cdots + (-1)^{t-1}\frac{(u-t)!}{(t-1)!0!}\right]$
2	$\frac{t!}{2!u!}\left[\frac{(u-2)!}{0!(t-2)!} - \frac{(u-3)!}{1!(t-3)!} + \cdots + (-1)^{t-2}\frac{(u-t)!}{(t-2)!0!}\right]$
...	...
t	$\frac{t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]$

We now determine the first four moments, ν_1, ν_2, ν_3, and ν_4, of this distribution about the origin $s = 0$. Since, in general,

$$\nu_k = \sum_{s=0}^{t} [s^k \times (\text{Relative frequency of } s)] = \sum_{s=0}^{t} s^k P_{(s)},$$

the tabulation for the computation of any moment is as follows:

Table II: The Computation of the k^{th} Moment of the Distribution of s

s	$s^k P_{(s)}$
0	0
1	$\frac{1^k t!}{1!u!}\left[\frac{(u-1)!}{0!(t-1)!} - \frac{(u-2)!}{1!(t-2)!} + \frac{(u-3)!}{2!(t-3)!} - \cdots + (-1)^{t-1}\frac{(u-t)!}{(t-1)!0!}\right]$
2	$\frac{2^k t!}{2!u!}\left[\frac{(u-2)!}{0!(t-2)!} - \frac{(u-3)!}{1!(t-3)!} + \cdots + (-1)^{t-2}\frac{(u-t)!}{(t-2)!0!}\right]$
3	$\frac{3^k t!}{3!u!}\left[\frac{(u-3)!}{0!(t-3)!} - \cdots + (-1)^{t-3}\frac{(u-2)!}{(t-3)!0!}\right]$
...	...
t	$\frac{t^k t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]$.

Noting that $\frac{1^k}{1!} = \frac{1^{k-1}}{0!}, \frac{2^k}{2!} = \frac{2^{k-1}}{1!}, \cdots \frac{t^k}{t!} = \frac{t^{k-1}}{(t-1)!}$, and multiplying the terms in brackets by these factors, we develop Table III:

Table III

s	$s^k P_{(s)}$
	1st diagonal ↓ 2nd diagonal ↓ 3rd diagonal ↓
0	0
1	$\frac{t!}{u!}\left[\frac{1^{k-1}(u-1)!}{0!0!(t-1)!} - \frac{1^{k-1}(u-2)!}{0!1!(t-2)!} + \frac{1^{k-1}(u-3)!}{0!2!(t-3)!}\right.$ $\left. - \cdots + (-1)^{t-1}\frac{1^{k-1}(u-t)!}{0!(t-1)!0!}\right]$ (t^{th} diagonal ↓) (t terms)
2	$\frac{t!}{u!}\left[\frac{2^{k-1}(u-2)!}{1!0!(t-2)!} - \frac{2^{k-1}(u-3)!}{1!1!(t-3)!}\right.$ $\left. + \cdots + (-1)^{t-2}\frac{2^{k-1}(u-t)!}{1!(t-2)!0!}\right]$ ($t-1$ terms)
3	$\frac{t!}{u!}\left[\frac{3^{k-1}(u-3)!}{2!0!(t-3)!} - \cdots + (-1)^{t-3}\frac{3^{k-1}(u-t)!}{2!(t-3)!0!}\right]$ ($t-2$ terms)
...	...
t	$\frac{t!}{u!}\left[\frac{t^{k-1}(u-t)!}{(t-1)!0!0!}\right]$ (1 term)

Since each series in brackets is one term shorter than the preceding series, the table forms a system of t diagonals. The sum which gives us ν_k may therefore be considered as the sum of these diagonals.

Now, from inspection, it is evident that the general diagonal is of the form

$$s^{\text{th}}\text{ diagonal} = \frac{t!(u-s)!}{u!(t-s)!}\left[\frac{s^{k-1}}{(s-1)!0!} - \frac{(s-1)^{k-1}}{(s-2)!1!} + \cdots + (-1)^{s-1}\frac{1^{k-1}}{0!(s-1)!}\right]$$

$$= \frac{t!(u-s)!}{u!(t-s)!}\left[\frac{1}{(s-1)!}\sum_{r=0}^{s-1}(-1)^r(s-r)^{k-1}\binom{s-1}{r}\right].$$

But it can be shown[4] that

$$\sum_{r=0}^{s-1}(-1)^r(s-r)^{k-1}\binom{s-1}{r} = 0 \qquad \text{when} \qquad k < s.$$

Whence

$$s^{\text{th}}\text{ diagonal} = 0 \qquad \text{when} \qquad k < s.$$

Therefore ν_k is given simply by the sum of the first k diagonals of Table III. Or, in general,

$$\nu_k = \frac{t!(u-1)!}{s!(t-1)!}\left[\frac{1^{k-1}}{0!0!}\right]$$

$$+ \frac{t!(u-2)!}{u!(t-2)!}\left[\frac{2^{k-1}}{1!0!} - \frac{1^{k-1}}{0!1!}\right]$$

$$+ \frac{t!(u-3)!}{u!(t-3)!}\left[\frac{3^{k-1}}{2!0!} - \frac{2^{k-1}}{1!1!} + \frac{1^{k-1}}{0!2!}\right]$$

$$+ \quad \cdots$$

$$+ \frac{t!(u-k)!}{u!(t-k)!}\left[\frac{k^{k-1}}{(k-1)!0!} - \frac{(k-1)^{k-1}}{(k-2)!1!} + \cdots + (-1)^{k-1}\frac{1^{k-1}}{0!(k-1!)}\right]. \qquad (3)$$

To this equation we must, of course, add the condition $k \leqq t$.

[4] E. Netto, *Lehrbuch der Combinatorik*, Leipzig, 1901, 249, Formula 17.

Solving now for the first four moments, we have

$$\left.\begin{aligned}
\nu_1 &= \frac{t}{u}, \\
\nu_2 &= \frac{t}{u}\left[1 + \frac{t-1}{u-1}\right], \\
\nu_3 &= \frac{t}{u}\left[1 + 3\frac{t-1}{u-1} + \frac{(t-1)(t-2)}{(u-1)(u-2)}\right], \\
\nu_4 &= \frac{t}{u}\left[1 + 7\frac{t-1}{u-1} + 6\frac{(t-1)(t-2)}{(u-1)(u-2)} + \frac{(t-1)(t-2)(t-3)}{(u-1)(u-2)(u-3)}\right].
\end{aligned}\right\} \quad (4)$$

If now we define, for convenience,

$$a = \frac{t}{u},$$

$$b = \frac{t-1}{u-1},$$

$$c = \frac{t-2}{u-2},$$

$$d = \frac{t-3}{u-3},$$

we have, for the constants of the distribution of s,

$$\left.\begin{aligned}
\text{Mean} &= \nu_1 = a. \\
\mu_2 &= \nu_2 - \nu_1^2 \\
&= a(1+b) - a^2, \quad \text{whence } \sigma = \sqrt{a(1+b) - a^2}. \\
\mu_3 &= \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3 \\
&= a(1 + 3b + bc) - 3a^2(1+b) + 2a^3 \\
\mu_4 &= \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4 \\
&= a(1 + 7b + 6bc + bcd) - 4a^2(1 + 3b + bc) + 6a^3(1+b) - 3a^4.
\end{aligned}\right\} \quad (5)$$

From these constants we can determine the skewness and kurtosis of the distribution of s,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}. \quad (6)$$

Now it is known that the means of samples of size n drawn from a parent population with constants β_1 and β_2 are distributed in such a way that

$$\beta_{1(\text{means})} = \frac{\beta_1}{n}, \qquad \text{and} \qquad \beta_{2(\text{means})} = 3 + \frac{\beta_2 - 3}{n}. \tag{7}$$

Therefore, having determined the beta-constants for the distribution of s, we can determine the beta-constants of the distribution of $\bar{s}$, the mean number of correct matchings resulting from n independent trials.
Now when $t = u \geqq 4$, we have

$$a = b = c = d = 1,$$

and equations (5) give us for the distribution of s

$$\left.\begin{aligned} \text{Mean} &= 1, \\ \mu_2 &= 1, \\ \mu_3 &= 1, \\ \mu_4 &= 4. \end{aligned}\right\} \quad \text{whence} \quad \begin{cases} \beta_1 = 1 \\ \beta_2 = 4 \end{cases}$$

and therefore, for the constants of the distribution of $\bar{s}$, we have, by equations (7),

$$\beta_1 = \frac{1}{n}, \qquad \text{and} \qquad \beta_2 = 3 + \frac{1}{n},$$

which indicates a positively skewed and leptokurtic distribution. The effect of increasing u and holding t constant is to increase the skewness, as shown in the following table for $t = 5$:

t	u	β_1
5	5	$\frac{1}{n}$
5	6	$\frac{1.05}{n}$
5	7	$\frac{1.16}{n}$
5	8	$\frac{1.31}{n}$
5	9	$\frac{1.46}{n}$

The degrees of skewness and kurtosis met with in practical cases of matching with any considerable number of judges (n) are such that a Pearson Type III distribution curve gives a reasonably good fit to the distribution of mean numbers of correct matchings. If, therefore, we have to determine the significance

of any obtained mean number of correct matchings, we may resort to Salvosa's tables[5] of the area under the Type III curve.

As a concrete example of the application of this method let us imagine that 10 judges have arranged 5 character sketches against 8 specimens of handwriting, 5 of which are true apposites of the sketches. Let the total number of correct matchings achieved by this group be 12, whence the mean number per judge is 1.2. We have, then,

$$\bar{s} = 1.2, \qquad n = 10,$$

$$t = 5, \qquad u = 8, \qquad \text{whence} \qquad a = \frac{t}{u} = .625,$$

$$b = \frac{t-1}{u-1} = .571,$$

$$c = \frac{t-2}{u-2} = .500.$$

We now find the mean, standard deviation, and β_1 of the distribution of $\bar{s}$, as follows:

The mean of the distribution of $\bar{s}$ is, by sampling theory, the same as the mean of the distribution of s:

$$\text{Mean} = a = .625.$$

The second moment of the distribution of $\bar{s}$ is, by sampling theory, $\frac{1}{n}$ times the second moment of the distribution of s; whence, by equation (5),

$$\text{Standard deviation} = \sqrt{\frac{1}{10}[a(1+b) - a^2]} = .243.$$

And, by equations (5) and (7),

$$\beta_1 = \frac{1}{10}\frac{[a(1+3b+bc) - 3a^2(1+b) + 2a^3]^2}{[a(1+b) - a^2]^3} = .131.$$

Now the obtained mean number of correct matchings was 1.2, and the next lower number which could have occurred (corresponding to a total of 11 instead of 12 for the group of judges) is 1.1. The lower boundary of the class-interval whose midpoint is $\bar{s} = 1.2$ is therefore 1.15; and it is the area above this boundary under the curve of $\bar{s}$ in which we are interested.

[5] L. R. Salvosa, Tables of Pearson's Type III function, *Ann. Math. Statist.*, 1, 1930, 191–198.

The deviation of this boundary from the mean of $\bar{s}$ is

$$1.15 - .625 = .525\,,$$

and this deviation expressed in terms of the standard deviation gives

$$\frac{.525}{.243} = 2.16\,.$$

Entering Salvosa's table for the deviation 2.16 and skewness $= \sqrt{\beta_1} = .36$, we find by interpolation that so good a performance should be expected by chance only about 23 times in 1000.

MOMENTS ABOUT THE ARITHMETIC MEAN OF A BINOMIAL FREQUENCY DISTRIBUTION

W. J. KIRKHAM, *Oregon State College*

Although the most useful moments of a binomial distribution have been derived as a function of the parameters of the generating binomial for any binomial frequency series, a generalization of notation and procedure is well worth our consideration. The problem attempted in this paper is the calculation of the moments about the mean for the general frequency series of Table I.

TABLE I

The generalized binomial frequency series

x (item)	f (frequency)
0	$N \cdot {}_nC_0 p^0 q^n$
1	$N \cdot {}_nC_1 p^1 q^{n-1}$
2	$N \cdot {}_nC_2 p^2 q^{n-2}$
.	.....
.	.....
n	$N \cdot {}_nC_n p^n q^0$

In calculating the moments of a set of data about any value, it is often found convenient to use an arbitrary origin, define the moments about this value, and represent the desired moments in terms of those calculated. In the general binomial series, the origin of the x's is found to be the best arbitrary origin. These intermediate moments are

$$\begin{aligned} \nu_1 &= \frac{\sum fx}{N} = M, \text{ arithmetic mean;} \\ \nu_2 &= \frac{\sum fx^2}{N}; \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ \nu_n &= \frac{\sum fx^n}{N} \end{aligned} \tag{1}$$

where ν_i is the i^{th} moment.

The moments (μ's) about the mean are easily defined as functions of the ν's

from fundamental definitions of the μ's. Denoting the i^{th} moment by μ_i, we have

$$\begin{aligned}
\mu_1 &= \frac{\sum f(x - M)}{N} = 0, \\
\mu_2 &= \frac{\sum f(x - M)^2}{N} = \nu_2 - \nu_1^2, \\
\mu_3 &= \frac{\sum f(x - M)^3}{N} = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{2}$$

In general,

$$\mu_n = \nu_n - {}_nC_1\nu_{n-1}\nu_1 + {}_nC_2\nu_{n-2}\nu_1^2 + \cdots + (-1)^{n-1}({}_nC_{n-1} - 1)\nu_1^n. \tag{3}$$

Or, if we let $\{\nu\}^n = \nu_n$, we may express the n^{th} moment by a simple notation.

$$\mu_n = \{\mu\}^n = \{\nu\}^n - {}_nC_1\{\nu\}^{n-1}\nu_1 + {}_nC_2\{\nu\}^{n-2}\nu_1^2 + \cdots = (\{\nu\} - \nu_1)^n. \tag{4}$$

Solving the equation for $\{\nu\}$,

$$\{\nu\} = \{\mu\} + \nu_1.$$

Raising both sides to the n^{th} power and substituting for the "brace" notation,

$$\nu_n = \mu_n + {}_nC_1\mu_{n-1}\nu_1 + {}_nC_2\mu_{n-2}\nu_1^2 + \cdots + \nu_1^n.$$

Whence

$$\mu_n = \nu_n - {}_nC_1\mu_{n-1}\nu_1 - {}_nC_2\mu_{n-2}\nu_1^2 - \cdots - \nu_1^n, \tag{5}$$

a semi-recursion formula.

The original formula for μ_n contained n moments or variables; and since there are only $(n - 2)$ of the μ's which are of lower order than μ_n, it is necessary to retain ν_n and ν_1 in (5). Since $\mu_1 = 0$, one term in the expansion of μ_n is zero. For instance, when $n = 5$, we have

$$\mu_5 = \nu_5 - 5\mu_4\nu_1 - 10\mu_3\nu_1^2 - 10\mu_2\nu_1^3 - \nu_1^5.$$

To calculate μ_k, it is necessary to calculate the ν's from ν_1 to ν_k. For the binomial series, these ν's are

$$\begin{aligned}
\nu_1 &= 1npq^{n-1} + \frac{2(n)(n-1)}{1\cdot 2}p^2q^{n-2} + \frac{3(n)(n-1)(n-2)}{1\cdot 2\cdot 3}p^3q^{n-3} + \cdots \\
&= np\left[q^{n-1} + \frac{(n-1)}{1}p^1q^{n-2} + \frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + p^{n-1}\right] \\
&= np(q + p)^{n-1} = np, \\
\nu_2 &= np\left[1\cdot q^{n-1} + \frac{2(n-1)}{1!}p^1q^{n-2} + 3\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + np^{n-1}\right],
\end{aligned}$$

$$\nu_3 = np\left[1^2 \cdot q^{n-1} + 2^2\frac{(n-1)}{1!}p^1q^{n-2} + 3^2\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^2p^{n-1}\right],$$

. . . . . .

. . . . . .

$$\nu_k = np\left[1^{k-1}q^{n-1} + 2^{k-1}\frac{(n-1)}{1!}p^1q^{n-2} + 3^{k-1}\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^{k-1}p^{n-1}\right].$$

In the simplified form of ν_k, the [ ] is the $(k-1)^{\text{th}}$ moment about -1 of the binomial series generated by the binomial $(q+p)^{n-1}$. Denoting this [ ] by $\nu'_{k-1}(n-1)$, the ν's can be expressed by the formula

$$\nu_k = np\nu'_{k-1}(n-1), \tag{6}$$

where ν' is a function of $(n-1)$ and $(k-1)$ while ν_k was a function of n and k.

Let us see how a ν' in ν_k can be defined in terms of the ν's of lower order than k. In finding this relationship, a consideration of the two series of Table II will be helpful.

TABLE II

x'	f	x	f
1	$N_{n-1}C_0p^0q^{n-1}$	0	$N_{n-1}C_0p^0q^{n-1}$
2	$N_{n-1}C_1p^1q^{n-2}$	1	$N_{n-1}C_1p^1q^{n-2}$
.	. . . . .	.	. . . . .
.	. . . . .	.	. . . . .
n	$N_{n-1}C_{n-1}p^nq^0$	$n-1$	$N_{n-1}C_{n-1}p^{n-1}q^0$

The [ ] in ν_k for Table I is equal to the $(k-1)^{\text{th}}$ moment of x' about $x' = 0$. Or

$$\nu_{k-1}, \text{ Table II}, x', = \nu_{k-1}, \text{ Table I}, = \nu'_{k-1}(n-1).$$

Also ν_{k-1} for x, Table II, is ν_{k-1} for the series generated by $(q-p)^{n-1}$.

The desired relationship between the ν's for the two series of Table II can be found by making use of the equations expressing the equality of the μ's for x and x'. Dropping the variable which shows the number of items, the same for the two series of Table II, in the notation, we have

$$\mu_2 = \mu'_2 = \nu_2 - \nu_1^2 = \nu'_2 - \nu'^2_2, \qquad \nu'_2 = \nu_2 - 2\nu_1(\nu_1 - \nu'_1) + (\nu_1 - \nu'_1)^2;$$

$$\mu_3 = \mu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = \nu'_3 - 3\nu'_2\nu'_1 + 2\nu'^3_1,$$

$$\nu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 + 3\nu'_2\nu'_1 - 2\nu'^3_1.$$

Substituting the value of ν'_2 in the right member of ν'_3,

$$\nu'_3 = \nu_3 - 3\nu_2(\nu_1 - \nu'_1) + 3\nu_1(\nu_1 - \nu'_1)^2 - (\nu_1 - \nu'_1)^3.$$

In general,

$$\nu_k' = \nu_k - {}_kC_1\nu_{k-1}(\nu_1 - \nu_1') + {}_kC_2\nu_{k-2}(\nu_1 - \nu_1')^2 + \cdots + (-1)^k(\nu_1 - \nu_1')^k. \quad (7)$$

The formula just derived may be used to define the moments about any origin in terms of those about the original zero of the x's. For our immediate use, the formula simplifies since $\nu_1' = \nu_1 + \nu_0 = \nu_1 + 1$. Then

$$\nu_k' = \nu_k + {}_kC_1\nu_{k-1} + {}_kC_2\nu_{k-2} + {}_kC_3\nu_{k-3} + \cdots \quad (8)$$

By simple analysis we found the value of ν_1 to be np. By the method of continuation, we are able to extend the list of ν's to any number. ν' from (8) is used in (6) with n replaced by $(n - 1)$ in the ν's.

$\nu_0 = 1.$

$\nu_1 = np.$

$\nu_2 = np\nu_1'(n - 1) = np[\nu_1(n - 1) + \nu_0(n - 1)]$

$= np[(n - 1)p + 1] = n(n - 1)p^2 + np.$

$\nu_3 = np\nu_2'(n - 1) = np[\nu_2(n - 1) + 2\nu_1(n - 1) + \nu_0(n - 1)]$

$= n(n - 1)(n - 2)p^3 + 3n(n - 1)p^2 + np.$

$\nu_4 = np\nu_3'(n - 1) = np[\nu_3(n - 1) + 3\nu_2(n - 1) + 3\nu_1(n - 1) + \nu_0(n - 1)]$

$= np\{[(n - 1)(n - 2)(n - 3)p^3 + 3(n - 1)(n - 2)p^2 + (n - 1)p]$

$+ 3[(n - 1)(n - 2)p^2 + (n - 1)p] + 3[(n - 1)p] + 1\}.$

$= n(n - 1)(n - 2)(n - 3)p^4 + 6(n)(n - 1)(n - 2)p^3 + 7(n)(n - 1)p^2 + np.$

. . . . . .

If the order of the terms in the expansion is reversed, ν_n is an ascending power series in p. The pure numerical coefficients in some of these ν's are

$\nu_1 = (1)$

$\nu_2 = (1, 1)$

$\nu_3 = (1, 3, 1)$

$\nu_4 = (1, 7, 6, 1)$

$\nu_5 = (1, 15, 25, 10, 1)$

$\nu_6 = (1, 31, 90, 65, 15, 1)$

$\nu_7 = (1, 63, 301, 350, 140, 21, 1)$

$\nu_8 = (1, 127, 966, 1701, 1050, 266, 28, 1).$

In general,

$$\nu_{n+1} = \left(1, \sum_{1}^{n} {}_nC_i, \sum_{2}^{n}\left({}_nC_i \sum_{1}^{i-1} {}_{i-1}C_j\right), \sum_{3}^{n}\left({}_nC_i \sum_{2}^{i-1}\left({}_{i-1}C_j \sum_{1}^{j-1} {}_{j-1}C_k\right)\right), \cdots\right). \tag{9}$$

Using the foregoing ν's, and the semi-recursion formula, we are able to determine the μ's.

$$\mu_2 = \nu_2 - \nu_1^2$$
$$= [np + (n)(n-1)p^2] - (np)^2$$
$$= np(1-p)$$
$$= npq.$$

$$\mu_3 = \nu_3 - 3\nu_1\mu_2 - \nu_1^3$$
$$= [np + 3n(n-1)p^2 + (n)(n-1)(n-2)p^3] - 3(np)[np(1-p)] - [np]^3.$$
$$= np + (-3n)p^2 + (2n)p^3 = np(1 - 3p + 2p^2)$$
$$= np(1-p)(1-2p)$$
$$= npq(q-p).$$

$$\mu_4 = [np + 7(n)(n-1)p^2 + 6(n)(n-1)(n-2)p^3 + (n)(n-1)(n-2)(n-3)p^4] - 4(np)(np)(1 - 3p + 2p^2) - 6(np)^2(np)(1-p) - (np)^4$$
$$= np + (-7n + 3n^2)p^2 + (12n - 6n^2)p^3 + (-6n + 3n^2)p^4$$
$$= np(1 - 7p + 12p^2 - 6p^3) + 3n^2p^2(1 - 2p + p^2)$$
$$= npq - 6np^2q^2 + 3n^2p^2q^2.$$

$$\mu_5 = np(1 - 15p + 50p^2 - 60p^3 + 24p^4) + 10n^2p^2(1 - 4p + 5p^2 - 2p^3)$$
$$= (q-p)(npq - 12np^2q^2 + 10n^2p^2q^2).$$

$$\mu_6 = np(1 - 31p + 180p^2 - 390p^3 + 360p^4 - 120p^5) + 5n^2p^2(5 - 36p + 83p^2 - 78p^3 + 26p^4) + 15n^3p^3(1 - 3p + 3p^2 - p^3)$$
$$= npq - 30np^2q^2(q-p)^2 + 25n^2p^2q^2 - 130n^2p^3q^3 + 15n^3p^3q^3.$$

$$\mu_7 = np(1 - 63p + 602p^2 - 2100p^3 + 3360p^4 - 2520p^5 + 720p^6) + n^2p^2(56 - 686p + 2590p^2 - 4270p^3 + 3234p^4 - 924p^5) + n^3p^3(105 - 525p + 945p^2 - 735p^3 + 210p^4)$$

$$= (q - p)(npq - 60np^2q^2 + 360np^3q^3 + 56n^2p^2q^2 - 462n^2p^3q^3 + 105n^3p^3q^3).$$

$$\mu_8 = np(1 - 127p + 1932p^2 - 10206p^3 + 25200p^4 - 31920p^5 + 20160p^6 - 5040p^7) + n^2p^2(119 - 2394p + 13895p^2 - 35700p^3 + 46004p^4 - 29232p^5 + 7308p^6) + n^3p^3(490 - 3850p + 10990p^2 - 14770p^3 + 9520p^4 - 2380p^5) + n^4p^4(105 - 420p + 630p^2 - 420p^3 + 105p^4)$$

$$= npq(1 - 42pq(3 - 40pq(1 - 3pq))) + 7n^2p^2q^2(17 - 4pq(77 - 261pq)) + 70n^3p^3q^3(7 - 34pq) + 105n^4p^4q^4.$$

ON CERTAIN DISTRIBUTION FUNCTIONS WHEN THE LAW OF THE UNIVERSE IS POISSON'S FIRST LAW OF ERROR[1]

By Frank M. Weida

Introduction. The median, which is that value of a permuted variable which has as many observed values on one side of it as on the other, appears to be the natural competitor of the arithmetic mean when we are interested in the probable or most probable value of an unknown quantity. It is well known[2] that the law of probability, namely, Poisson's first law of error, which results from the assumption that the median is the most probable value of the unknown quantity is given by

$$f(x) = \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}}. \tag{1}$$

Little is known about the form of the distribution functions of the more important statistics when the law of the "Universe" is Poisson's first law of error. It, therefore, appears to be of interest and importance to enlarge our present knowledge of distribution functions by finding certain new ones when the variable or variables are defined by (1).

In this paper we present the following results: (1) We have obtained an explicit expression for the distribution of means of samples of n; (2) we have obtained an explicit expression for the distribution of differences; (3) we have obtained an explicit expression for the distribution of quotients; (4) we have obtained an explicit expression for the distribution of standard deviations for samples of n; (5) we have obtained an explicit expression for the distribution of geometric means for samples of n; (6) we have obtained an explicit expression for the distribution of harmonic means for samples of n.

In our analysis, we have made use of the theory of characteristic functions in the sense of Levy.[3] This theory has been extended to more than one dimension by V. Romanovsky[4] and by E. K. Haviland.[5] S. Kullback,[6] in his thesis, has made further extensions and has applied them successfully to the distribution problem in statistics.

[1] Presented to the American Mathematical Society, February 23, 1935.

[2] Brunt, David: "The Combination of Observations," 1923, p. 27.

[3] Levy, P.: "Calcul des Probabilités;" pp. 153–191.

[4] Romanovsky, V.: "Sur un théorème limite du calcul des probabilités," Recueil mathématique de la Société mathématique de Moscow, Vol. 36, 1926, pp. 36–64.

[5] Haviland, E. K.: "On the inversion formula for Fourier-Stieltjes transforms in more than one dimension," American Journal of Mathematics, Vol. 57, 1935, pp. 94–101.

[6] Kullback, S.: "An application of characteristic functions to the distribution problem of statistics," Annals of Mathematical Statistics, Vol. V, No. 4, pp. 263–307.

The explicit expression for the distribution of arithmetic means of samples of n is not new. This law of distribution has previously been obtained otherwise by F. Hausdorff[7] and by A. T. Craig.[8] It is inserted here to show the superiority and greater power of our method when compared with previous methods and for the completeness of our discussion. The other results offered in this paper, as far as the writer knows, are new.

1. **The distribution of arithmetic means.** Let us consider

$$f(x) = \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}}, \qquad (-\alpha < x < \alpha). \tag{2}$$

If we assume that $x_1, x_2, \cdots, x_n$ are independently distributed and that each $x_i (i = 1, 2, \cdots, n)$ is distributed according to the same distribution law, namely, Poisson's first law of error, then it is fairly easy to see that the characteristic function for the law of distribution of means of samples of n is given by

$$\phi(t) = \left\{ \int_{-\alpha}^{\alpha} \frac{k}{\sigma} e^{it|x| - \frac{|x|}{\sigma}} \, dx \right\}^n. \tag{3}$$

If $u = \Sigma_i x_i$ $(i = 1, 2, \cdots, n)$, then it follows that the distribution function of u, namely, $P(u)$, is given by

$$P(u) = \frac{1}{2\pi} \int_{-\alpha}^{\alpha} e^{-itu} \left\{ \frac{2k}{\sigma} \int_0^{\alpha} e^{itx - \frac{x}{\sigma}} \, dx \right\}^n dt, \tag{4}$$

which, upon simplification becomes

$$P(u) = \frac{2^{n-1} k^n}{\pi \sigma^n} \int_{-\alpha}^{\alpha} \frac{e^{-itu} \, dt}{(1 - \sigma it)^n}. \tag{5}$$

It is readily seen that the poles of the integrand are of the n^{th} order and are those of $(1 - \sigma it)^n$. It follows by the well known Residue Theorem of Cauchy[9] that

$$P(u) = \frac{2^{n-1} k^n}{\pi \sigma^n} \cdot 2\pi i \cdot \frac{(-1)^{n-1}}{(n-1)!} \cdot \frac{1}{i^n} \cdot \frac{d^{n-1}}{dt^{n-1}} \left\{ \frac{e^{-itu}}{(1 + \sigma it)^n} \right\}_{t = \frac{1}{\sigma i}}. \tag{6}$$

If now, we replace u by $n|\bar{x}|$, we will obtain the desired law of the distribution of arithmetic means of samples of n which is

$$P(|\bar{x}|) = \frac{2^n k^n (-1)^{n-1} n}{\sigma^n i^{n-1} (n-1)!} \cdot \frac{d^{n-1}}{dt^{n-1}} \left\{ \frac{e^{-itn|\bar{x}|}}{(1 + \sigma it)^n} \right\}_{t = \frac{1}{\sigma i}} \tag{7}$$

defined for all values of x on the range $(-\alpha < x < \alpha)$.

---

[7] Hausdorff, F.: Beitrage zur Wahrscheinlichkeitsrechnung Königlich Sächsischen Gesellschaft der Wissenshaften zu Leipzig. Berichted über die Verhandlungen Math.-Phys. Classe, Vol. 53, 1901, pp. 152–178.

[8] Craig, A. T.: "On the distribution of certain statistics," American Journal of Mathematics, Vol. 54, 1932, pp. 353–366.

[9] Macrobert, T. M.: "Functions of a Complex Variable," 1933, pp. 57, 295.

A. T. Craig[8] has given the distribution laws of arithmetic means of samples of size 2, 3, and 4. These results as well as the results for any n are readily obtained from (7).

2. **The distribution of differences.** Let us assume that the laws of distribution of x and y are independent and that they are given respectively by

$$f(x) = \frac{k_1}{\sigma_1} e^{-\frac{|x|}{\sigma_1}}; \qquad f(y) = \frac{k_2}{\sigma_2} e^{-\frac{|y|}{\sigma_2}}; \qquad (-\alpha < x < \alpha), \qquad (-\alpha < y < \alpha).$$

In this case, the characteristic function of the law of distribution of differences $(x - y)$ is given by

$$\phi(t) = \frac{k_1}{\sigma_1} \int_{-\alpha}^{\alpha} e^{it|x| - \frac{|x|}{\sigma_1}} dx \cdot \frac{k_2}{\sigma_2} \int_{-\alpha}^{\alpha} e^{-it|y| - \frac{|y|}{\sigma_2}} dy. \tag{8}$$

Performing the operations indicated in (8) and simplifying, we find that

$$\phi(t) = \frac{4k_1k_2}{\sigma_1\sigma_2} \cdot \frac{1}{(1 - \sigma_1 it)} \cdot \frac{1}{(1 + \sigma_2 it)}. \tag{9}$$

It is fairly easy to see that the distribution law of u is given by

$$P(u) = \frac{4k_1k_2}{2\pi\sigma_1\sigma_2} \int_{-\alpha}^{\alpha} \frac{e^{-itu}\, dt}{(1 - \sigma_1 it)(1 + \sigma_2 it)}. \tag{10}$$

Now, let $\{(1/\sigma_1) - it\} = v/u$, then (10) becomes

$$P(u) = \frac{2k_1k_2 e^{-\frac{u}{\sigma_1}}}{\pi i\sigma_1\sigma_2(\sigma_1 + \sigma_2)} \int_{-\frac{u}{\sigma_1} - i\alpha}^{-\frac{u}{\sigma_1} + i\alpha} \frac{e^{-v}\, dv}{(-v)\left\{1 + \frac{v}{\left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)}\right\}}. \tag{11}$$

The integral in (11) is convergent because

$$\lim_{v \to \alpha} \left| v^m \frac{e^{-v}\, dv}{(-v)\left\{1 + \frac{v}{\left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)}\right\}} \right| = 0.$$

Hence, we find that

$$P(u) = -\frac{2k_1k_2 e^{-\frac{u}{\sigma_1}}}{\pi i\sigma_1\sigma_2(\sigma_1 + \sigma_2)} \int_{\alpha}^{(0+)} \frac{e^{-v}\, dv}{(-v)\left\{1 + \frac{v}{\left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)}\right\}} \tag{12}$$

which upon simplification becomes

$$P(u) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1+\sigma_2)} W_{0,\frac{1}{2}}\left\{\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}u\right\}, \tag{13}$$

where $W_{0,\frac{1}{2}}\left\{\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}u\right\}$ is the *confluent hypergeometric function.*[10]

It is well known that

$$W_{k,m}(z) = \frac{e^{-\frac{1}{2}z}z^k}{\Gamma(\frac{1}{2}-k+m)}\int_0^\alpha t^{-k-\frac{1}{2}+m}\left(1+\frac{t}{z}\right)^{k-\frac{1}{2}+m}e^{-t}\,dt$$

for all values of k and m and for all values of z except negative real values. Clearly,

$$W_{0,\frac{1}{2}}\left\{\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}u\right\} = \frac{e^{-\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}u}}{\Gamma(1)}\int_0^\alpha e^{-t}\,dt$$

which, upon simplification becomes

$$W_{0,\frac{1}{2}}\left\{\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}u\right\} = e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}u}. \tag{14}$$

Hence, we now find that

$$P(u) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1+\sigma_2)}e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}u}. \tag{15}$$

If now, we replace u by $|x| - |y|$, we will obtain the desired law of distribution of differences which is

$$P(|x| - |y|) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1+\sigma_2)}e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}\{|x|-|y|\}}. \tag{16}$$

3. **The distribution of ratios.** We assume that the laws of distribution of x and y are independent and that they are given respectively by

$$f(x) = \frac{k_1}{\sigma_1}e^{-\frac{|x|}{\sigma_1}}; \quad f(y) = \frac{k_2}{\sigma_2}e^{-\frac{|y|}{\sigma_2}}; \quad (-\alpha < x < \alpha), \quad (-\alpha < y < \alpha).$$

[10] Whittaker, E. T. and Watson, G. N.: "A course in modern Analysis," 1915, pp. 333–334.

Let $u = \log |x| - \log |y|$. The characteristic function of the law of distribution of quotients is then given by

$$\phi(t) = \frac{k_1}{\sigma_1}\int_{-\alpha}^{\alpha} e^{-\frac{|x|}{\sigma_1}}(|x|)^{it}\,dx \cdot \frac{k_2}{\sigma_2}\int_{-\alpha}^{\alpha} e^{-\frac{|y|}{\sigma_2}}(|y|)^{-it}\,dy$$
$$= \frac{4k_1k_2}{\sigma_1\sigma_2}\int_0^{\alpha} e^{-\frac{x}{\sigma_1}}x^{it}\,dx \int_0^{\alpha} e^{-\frac{y}{\sigma_2}}y^{-it}\,dy. \tag{17}$$

Now, let $s = x/\sigma_1$ and $w = y/\sigma_2$, then clearly

$$\phi(t) = 4k_1k_2\sigma_1^{it}\sigma_2^{-it}\int_0^{\alpha} e^{-s}s^{it}\,ds \int_0^{\alpha} e^{-w}w^{-it}\,dw,$$

whence

$$\phi(t) = 4k_1k_2\sigma_1^{it}\sigma_2^{-it}\,\Gamma(it+1)\,\Gamma(1-it). \tag{18}$$

It follows that the distribution law of u is given by

$$P(u) = \frac{4k_1k_2}{2\pi}\int_{-\alpha}^{\alpha} e^{-itu+i\log\sigma_1 t - i\log\sigma_2 t}\,\Gamma(it+1)\,\Gamma(1-it)\,dt$$

which upon simplification, becomes

$$P(u) = \frac{2k_1k_2}{\pi}\int_{-\alpha}^{\alpha} e^{-i(u-\log\sigma_1+\log\sigma_2)t}\,\Gamma(it+1)\,\Gamma(1-it)\,dt. \tag{19}$$

Now, let $(1-it) = -v$, then (19) becomes

$$P(u) = \frac{4k_1k_2}{2\pi i}\int_{-1-i\alpha}^{-1+i\alpha} e^{-v\{u-\log\sigma_1+\log\sigma_2\}-\{u-\log\sigma_1+\log\sigma_2\}}\,\Gamma(2+v)\,\Gamma(-v)\,dv. \tag{20}$$

Since it can be shown that[11]

$$(1/2\pi i)\int_{-1-i\alpha}^{-1+i\alpha} e^{-vu}\Gamma(2+v)\,\Gamma(-v)\,dv = \Gamma(2)\{1+(1/e^u)\}^{-2},$$

we find that (20) becomes

$$P(u) = \frac{4k_1k_2e^{-u}\sigma_1}{\sigma_2}\,\Gamma(2)\left\{1+\frac{\sigma_1}{\sigma_2e^u}\right\}^{-2}. \tag{21}$$

Now, put $e^u = |x|/|y| = R$, whence from (21) we will obtain the desired law of distribution of quotients which is

$$P(R) = \frac{4k_1k_2\sigma_1\Gamma(2)}{\sigma_2R}\left\{1+\frac{\sigma_1}{\sigma_2R}\right\}^{-2}. \tag{22}$$

[11] Macrobert, T. M., "Functions of a Complex Variable," 1933, pp. 114, 139, 151. Whittaker, E. T. and Watson, G. N., "A course in modern Analysis," 1915, pp. 283.

4. **The distribution of variances and standard deviations.** If we assume that the variance and standard deviation are calculated about a sample mean and if we let $u = \sum_{i=1}^{n-1} x_i^2$, and if the x_i are independently distributed and each x_i is distributed according to the same distribution law, namely, Poisson's first law of error, then it is clear that the characteristic function for the law of distribution of variances of samples of n is

$$\phi(t) = \left\{\frac{k}{\sigma}\int_{-\alpha}^{\alpha} e^{-\frac{|x|}{\sigma}+itx^2}\,dx\right\}^{n-1} = \left\{\frac{2k}{\sigma}\int_{0}^{\alpha} e^{itx^2-\frac{x}{\sigma}}\,dx\right\}^{n-1}. \tag{23}$$

Let I represent the integral in the right-hand member of (23). We obtain that $(dI/d\sigma) = I/\sigma^2$, whence $I = Ce^{-\frac{1}{\sigma}}$. Making use of the conditions:

$$\sigma \to \alpha, \qquad I \to \int_0^{\alpha} e^{itx^2}\,dx = e^{\frac{1}{4}\pi i}\frac{\sqrt{\pi}}{\sqrt{t}},$$

$\sigma \to \alpha$, $Ce^{-\frac{1}{\sigma}} \to C$, whence we find that

$$\int_0^{\alpha} e^{itx^2-\frac{x}{\sigma}}\,dx = e^{\frac{1}{4}\pi i}\frac{\sqrt{\pi}}{\sqrt{t}}\,e^{-\frac{1}{\sigma}}.$$

Clearly, it follows that

$$\phi(t) = \frac{2^{n-1}k^{n-1}e^{\frac{(n-1)\pi i}{4}}\pi^{\frac{n-1}{2}}}{\sigma^{n-1}t^{\frac{n-1}{2}}}\,e^{-\frac{n-1}{\sigma}}. \tag{24}$$

We now find that the distribution law of u is given by

$$P(u) = \frac{2^{n-1}k^{n-1}e^{\frac{(n-1)\pi i}{4}}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}}{2\pi\sigma^{n-1}}\int_{-\alpha}^{\alpha}\frac{e^{-itu}}{t^{\frac{n-1}{2}}}\,dt. \tag{25}$$

Evaluating the integral in (25) with a suitably chosen contour,[12] we find that

$$P(u) = \frac{2^{n-1}k^{n-1}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}2\pi}{2\pi\sigma^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}\,u^{\frac{n-3}{2}}e^{-u}. \tag{26}$$

Now, let $u = \sum_{i=1}^{n} x_i^2 = ns^2$, whence from (26) we will obtain the desired law of distribution of variances which is

$$P(s^2) = \frac{2^{n-1}k^{n-1}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}}{\sigma^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}\,n^{\frac{n-3}{2}}(s^2)^{\frac{n-3}{2}}e^{-s^2}. \tag{27}$$

[12] Macrobert, T. M., "Functions of a Complex Variable," 1933, p. 67.

The law of distributions of standard deviations can be obtained at once from (27) since $d(s^2) = 2s\,ds$.

We shall now give the specific laws of distribution of variances for samples of size 1, 2, 3, 4, and 5 when the law of the "Universe" is Poisson's first law of error. From (27),

For $n = 1$,

$$P(s^2) = 0, \qquad (0 < s^2 < \infty). \tag{28}$$

For $n = 2$,

$$P(s^2) = \frac{2^{\frac{1}{2}} k e^{-\frac{1}{\sigma}} e^{-s^2}}{\sigma s}, \qquad (0 < s^2 < \infty). \tag{29}$$

For $n = 3$,

$$P(s^2) = \frac{4k^2\pi e^{-\frac{2}{\sigma}} e^{-s^2}}{\sigma^2}, \qquad (0 < s^2 < \infty). \tag{30}$$

For $n = 4$,

$$P(s^2) = \frac{32k^3\pi e^{-\frac{3}{\sigma}} s e^{-s^2}}{\sigma^3}, \qquad (0 < s^2 < \infty). \tag{31}$$

For $n = 5$,

$$P(s^2) = \frac{80k^4\pi^2 e^{-\frac{4}{\sigma}} s^2 e^{-s^2}}{\sigma^4}, \qquad (0 < s^2 < \infty). \tag{32}$$

5. **The distribution of geometric means.** As before, we assume that the x_i are independently distributed and each x_i is distributed according to the same distribution law, namely, Poisson's first law of error. Then, clearly, the characteristic function for the law of distribution of geometric means of samples of n is

$$\phi(t) = \left\{\int_{-\alpha}^{\alpha} \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}} |x|^{it}\, dx\right\}^n = \left\{\frac{2k}{\sigma}\int_0^{\alpha} e^{-\frac{x}{\sigma}} x^{it}\, dx\right\}^n. \tag{33}$$

Now, put $s = x/\sigma$, then (33) becomes

$$\phi(t) = \left\{2k\sigma^{it}\int_0^{\alpha} e^{-s} s^{it}\, ds\right\}^n = 2^n k^n \sigma^{nit}\{\Gamma(it+1)\}^n. \tag{34}$$

It follows at once that the distribution law of u is

$$P(u) = \frac{2^n k^n}{2\pi}\int_{-\alpha}^{\alpha} e^{-i(u+n\log\sigma)t}\{\Gamma(it+1)\}^n\, dt. \tag{35}$$

Now, let $it + 1 = -v$, then (35) becomes

$$P(u) = \frac{-2^n k^n}{2\pi i} e^{u+n\log\sigma} \int_{-1-i\alpha}^{-1+i\alpha} e^{v(u+n\log\sigma)} \{\Gamma(-v)\}^n \, dv. \tag{36}$$

It is well known that (10)

$$\{\Gamma(-v)\}^n = \frac{(-1)^n \pi^n}{\sin^n \pi v \{\Gamma(v+1)\}^n}. \tag{37}$$

Using (37) in (36), we readily find that

$$P(u) = \frac{-2^n k^n}{2\pi i} e^{u+n\log\sigma} \int_{-i-i\alpha}^{-1+i\alpha} \frac{e^{v(u+n\log\sigma)}(-1)^n \pi^n}{\{\Gamma(v+1)\}^n \sin^n \pi v} \, dv. \tag{38}$$

It is fairly easy to see that the poles of the integrand in (38) are the poles of $\{\Gamma(-v)\}^n$ and that these poles are of the n^{th} order. Applying the well known Residue Theorem of Cauchy (8), we find that

$$P(u) = 2^n k^n e^{u+n\log\sigma} \sum_{a=0}^{\alpha} \frac{(-1)^{n+na+1}}{(n-1)!} \left\{ \frac{d^{n-1}}{dv^{n-1}} \left[\frac{e^{v(u+n\log\sigma)}}{\{\Gamma(v+1)\}^n} \right] \right\}_{v=a}. \tag{39}$$

Now, since $u = \log|x_1| + \log|x_2| + \cdots + \log|x_n|$, then clearly, the distribution law of the geometric mean, G, is obtained from the law of distribution for u by means of the transformation

$$u = \log (G)^n.$$

Hence, from (39), we find the desired law of distribution of geometric means of samples of n which is

$$P\{G\} = \frac{2^n k^n G^n \sigma^n}{\Gamma(n)} \sum_{a=0}^{\alpha} (-1)^{n+na+1} \left\{ \frac{d^{n-1}}{dv^{n-1}} \left[\frac{G^{nv}\sigma^{nv}}{\{\Gamma(v+1)\}^n} \right] \right\}_{v=a}. \tag{40}$$

6. **The distribution of harmonic means.** Let us assume that $f(x)$ is the law of distribution for x. It is well known[13] that the law of distribution of $x' = 1/x$ is given by

$$F(x') = (1/x'^2) f(1/x')$$

if $1/x$ is continuous on the range of definition of $f(x)$. Now, in case $f(x)$ is Poisson's first law of error, we find that

$$F(x') = F(1/x) = \frac{k}{\sigma} x^2 e^{-\frac{|x|}{\sigma}}; \qquad (-\alpha \leqq x < 0), \quad (0 < x \leqq \alpha). \tag{41}$$

---

[13] Dodd, E. L., "The frequency law of a function of one variable," Bulletin of the American Mathematical Society, Vol. 31, 1925, p. 28; "The frequency law of a function of variables with given frequency laws," Annals of Mathematics, Second Series, Vol. 27, 1925–26, p. 18.

We assume that the x_i' are independently distributed and each x_i' is distributed according to the same law of distribution, whence we find that the characteristic function for the law of distribution of harmonic means of samples of n is

$$\phi(t) = \left\{ \int_0^\alpha \frac{k}{\sigma} e^{it|x| - \frac{|x|}{\sigma}} x^2 \, dx \right\}^n, \tag{42}$$

from which, after simplification, we find that

$$\phi(t) = \frac{k^n 2^n \sigma^{2n}}{(1 - \sigma it)^{3n}}. \tag{43}$$

We now find that the law of distribution for u is

$$P(u) = \frac{2^n k^n \sigma^{2n}}{2\pi} \int_{-\alpha}^{\alpha} \frac{e^{-itu}}{(1 - \sigma it)^{3n}} \, dt,$$

which, after evaluation and simplification, becomes

$$P(u) = \frac{2^n k^n}{\sigma^n \Gamma(3n)} u^{3n-1} e^{-\frac{u}{\sigma}}. \tag{44}$$

Recalling that in this case, $u = 1/|x_1| + 1/|x_2| + \cdots + 1/|x_n|$, we make the transformation $u = n/H$, where H is the harmonic mean; whence, from (44), we find that the desired law of distribution of harmonic means of samples of n is given by

$$P(H) = \frac{2^n k^n n^{3n-1}}{\sigma^n \Gamma(3n)} \cdot H^{1-3n} e^{-\frac{n}{\sigma H}}. \tag{45}$$

7. **Conclusions.** We have shown that the same analysis is applicable to find the explicit expression for all the distribution laws we have discussed in this paper.

THE GEORGE WASHINGTON UNIVERSITY,
WASHINGTON, D. C.

ERRATA

In my paper* there appear two blunders which were called to my attention by A. T. Craig.

In section 4, pages 107–108, headed "The distribution of variances and standard deviations," I have obtained the distribution function of the sum of the squares of $n - 1$ independent values of x and not the distribution function of the sum of the squares of the deviations from the sample mean of the n independent values of x.

In section 2, pages 104–105, headed "The distribution of differences," I have obtained the distribution function of the differences of *absolute values* and not the distribution function of the actual differences.

---

* Weida, F. M., "On Certain Distribution Functions when the Law of the Universe is Poisson's First Law of Error," Annals of Mathematical Statistics, Vol. VI, No. 2, June, 1935, pp. 102-110.

ON THE PROBLEM OF CONFIDENCE INTERVALS

By J. Neyman

When discussing my paper read before the Royal Statistical Society on 19th June, 1934, Professor Fisher said that the extension of his work concerning the fiducial argument to the case of discontinuous distributions, as presented in my paper, has been reached at a great expense: that instead of exact probability statements we get only statements in the form of inequalities.

This remark raises the question whether the disadvantage of the solution which he mentioned (the inequalities instead of equalities) results from the unsatisfactory method of approach, or whether it is connected with the nature of the problem itself.

I think that the problem is of considerable general interest. For instance it may be asked whether the confidence intervals for the binomial distribution recently published by E. S. Pearson and C. J. Clopper,[1] which correspond to the probability statements in inequalities, could be bettered.

The purpose of the present note is to show, (1) that in some exceptional cases the exact probability solution of the problem exists and that then it may easily be found by the method described in Note I of my paper;[2] (2) that in the general case of discontinuous distribution exact probability statements in the problem of confidence intervals are impossible.

In particular it will be seen that exact probability statements are impossible in the case of the binomial distribution and so that the system of confidence intervals published by Clopper and Pearson could not be bettered.

In order to avoid any possible misunderstanding I shall start by restating the problem.

We shall consider a random discontinuous variate x, capable of having one or another of a finite, or at most denumerable set of values

$$x_1, x_2, \cdots x_n, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

We shall assume that the frequency function, say $p(x \mid \theta)$, of x depends upon one parameter θ, the value of which is unknown. The problem of confidence intervals consists in ascribing to every possible value of x e.g. to x_n, $(n = 1, 2, \cdots)$ a "confidence interval," say $\theta_1(n)$ to $\theta_2(n)$ such that the probability, P, of our being correct in stating

$$\theta_1(n) \leqq \theta \leqq \theta_2(n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

whenever we observe $x = x_n$ $(n = 1, 2, \cdots)$, is either:

---

[1] E. S. Pearson and C. J. Clopper: The Use of Confidence or Fiducial Limits in the Case of the Binomial. Biometrika Vol. XXVI, pp. 404–413.

[2] J. R. S. S. Vol. 97, p. 589.

(a) equal to a given value $\alpha < 1$ chosen in advance, or

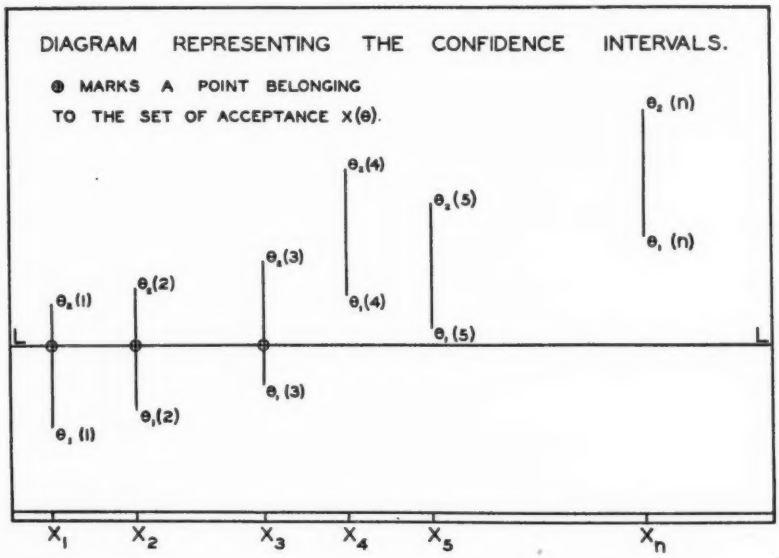
(b) *at least* equal to this value α.

I proposed to call this chosen value α the confidence coefficient.

In the earlier paper I showed that the solution of the problem in its form (b) is always possible and easy to find. If the variate x is continuous, then the solution of the problem (a) is equally easy. At present we shall consider whether and under what conditions the solution (a) is possible when the variate x is discontinuous.

Suppose that the variate x is discontinuous as described above, and that the solution of the problem in its form (a) exists and is given by the system of confidence intervals $(\theta_1(x_n), \theta_2(x_n))$ for $n = 1, 2, \cdots$.

The position is illustrated in the diagram below. On the axis of abscissae the possible values of the variate x are marked. The axis of ordinates is the axis of θ. The confidence intervals are marked on verticals passing through corresponding values of x.



According to our hypothesis the intervals $(\theta_1(x_n), \theta_2(x_n))$ are so chosen that

$$P = \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

P is the probability of an event, say E, which we shall describe in some detail. Let us denote generally the probability of any event a by $P\{a\}$. $P\{a \mid b\}$ will denote the probability of an event, a, calculated under the assumption that another event, b, has already occurred.

Now

$P = P\{E\}$ = the probability that {either $(x = x_1)$ and then $\theta_1(1) \leqq \theta \leqq \theta_2(1)$

or $(x = x_2)$ and then $\theta_1(2) \leqq \theta \leqq \theta_2(2)$

. . . . . .

. . . . . .

or $(x = x_n)$ " " $\theta_1(n) \leqq \theta \leqq \theta_2(n)$

. . . . . .

. . . . . .}

$$= P\{x = x_1\}P\{\theta_1(1) \leqq \theta \leqq \theta_2(1) \mid (x = x_1)\}$$

$$+ P\{x = x_2\}P\{\theta_1(2) \leqq \theta \leqq \theta_2(2) \mid (x = x_2)\}$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$= \sum_{n=1}^{\infty} P\{x = x_n\}P\{\theta_1(n) \leqq \theta \leqq \theta_2(n) \mid (x = x_n)\} = \alpha. \quad (4)$$

The calculation of the probability P in the above form is not convenient, as both multipliers in each term of the sum in (4) depend upon the unknown probability function *a priori* of θ. Therefore we shall present P in another form, giving to the event E a geometrical interpretation. Let us denote by CB the set of all confidence intervals $(\theta_1(n), \theta_2(n))$, as marked on the plane of x and θ. Thus CB will be composed of points with co-ordinates x and θ, where

$$\left.\begin{array}{c} x = x_n \\ \theta_1(n) \leqq \theta \leqq \theta_2(n) \end{array}\right\} n = 1, 2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

The set CB will be called the confidence belt.

Denote by A any point of the plane of x and θ, having any values for its co-ordinates.

It is easily seen that the event, which we denote by E, and the probability of which is $P = \alpha$, consists in the point A belonging to the confidence belt CB. In fact the event E occurs if and only if the co-ordinates of A fulfil the conditions (5). But just these conditions define the points belonging to CB.

The above circumstance allows us to calculate P by means of a formula which discloses its connection with $p(x \mid \theta)$.

Fix any possible value of $\theta = \theta'$ and draw the straight line LL the points of which have just this fixed value θ' for their ordinates. The line LL will cut some of the confidence intervals. Denote by $X(\theta')$ the set of points of intersection, and by $\phi(\theta)$ the unknown frequency function of θ. The set $X(\theta)$ will be called the set of acceptance corresponding to the specified value of θ.

The function $\phi(\theta)$ may be continuous or not. So may be $p(x \mid \theta)$ considered as a function of θ. These cases may be treated together if we agree that $\sum_{\theta} F(\theta)$ will denote either the sum or the integral of $F(\theta)$ extending over all values of θ, whenever $F(\theta)$ is integrable.

Using this notation we may write

$$P = P\{E\} = \sum_{\theta} \left\{ \phi(\theta) \sum_{X(\theta)} (p(x \mid \theta)) \right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

where $\sum_{X(\theta)}$ denotes the summation over all values of x belonging to $X(\theta)$.

From the formula (6) may be deduced the following important proposition.

The probability P may possess a constant value α, independent of the properties of the unknown function $\phi(\theta)$, if and only if for each θ

$$\sum_{X(\theta)} (p(x \mid \theta)) = \alpha. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

The condition (7) is obviously sufficient to have $P = \alpha$. In fact, if it is satisfied, then we should get from (6)

$$P = \alpha \sum_{\theta} (\phi(\theta)) = \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

since $\sum_{\theta} (\phi(\theta)) = 1$ whatever the frequency distribution of θ. It is equally easy to see that the condition (7) is necessary for having $P = \alpha$ whatever the function $\phi(\theta)$. For suppose that for $\theta = \theta_1$ we have

$$\sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Then if it happens, that

$$\phi(\theta_1) = 1 \qquad \text{for } \theta = \theta_1 \qquad (10)$$

and

$$\phi(\theta) = 0 \qquad \text{for } \theta \neq \theta_1 \qquad (11)$$

the only term in the sum $\sum_{\theta}$ which is different from zero will be that corresponding to $\theta = \theta_1$ and the formula (6) will reduce to

$$P = \sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha. \qquad (12)$$

The original question, whether the solution of the form (a) is possible when the variate x is discontinuous is thus put in the following form: is it possible to define for every possible value of θ a set of acceptance $X(\theta)$ such that the equation (7) holds good?

The answer is: in some cases it may be possible, but this depends upon the nature of the function $p(x \mid \theta)$. It is very easy to *invent* functions $p(x \mid \theta)$ for which the equation (7) for a definite value of α holds good, and we may even fix in advance the sets of acceptance $X(\theta)$. However the important question is not whether there may exist elaborately invented cases of discontinuous distributions where the solution (a) exists, but rather whether this solution exists always, or at least whether it exists frequently and in cases which are practically important.

This question must be answered in the negative on the basis of the following example concerning the most important of the discontinuous distributions, the Binomial.

In fact it will be seen below that if x is a variate following the binomial frequency law, then whatever the arrangement of the sets of acceptance $X(\theta)$, corresponding to different values of θ, the left hand side of the equation (7) cannot be constantly equal to the confidence coefficient $\alpha < 1$. It will follow that in the case of the binomial distribution, the solution of the problem (a) is impossible.

To prove this we shall consider the variate, x, following the binomial frequency law. That is to say we shall assume that x may have values 0, 1, 2, $\cdots n$, and that

$$p(x \mid \theta) = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{(n-x)} \tag{13}$$

while $0 < \theta < 1$. Since the set of possible values which x may have is finite, therefore the set of all confidence intervals must be finite also. It follows that there is possible only a finite number of sets of acceptance $X(\theta)$. Therefore there must be at least one set of acceptance, say X^0, which will be common to an infinite number of values of θ, say $\theta_1, \theta_2, \cdots \theta_n, \cdots$ so that for each it will be $X(\theta_n) = X^0$.

Now

$$\sum_{X(\theta_n)} (p(x \mid \theta_n)) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

for all these values of $\theta = \theta_n$ will be *the same* polynomial in θ of the order n. If it has the same value α for a number of values of θ exceeding n, it means that this polynomial is an absolute constant. Therefore if it were possible to give a solution of the type (a) in the case of the binomial distribution, it would be possible to construct a sum (14), the terms of which are all different and have the form (13), and such that after all possible reductions and simplifications all terms involving θ would cancel and we should be left only with one constant term $\alpha < 1$. This, however, is impossible, since the only term of the form (13) which involves a constant, is the term corresponding to $x = 0$

$$p(0 \mid \theta) = (1-\theta)^n = 1 - n\theta + \frac{n(n-1)}{2}\theta^2 \quad \ldots\ldots\ldots\ldots (15)$$

and then this constant is 1. Other terms of the form (13) involve θ^x as a multiplier. Therefore there exists only one sum of the form (14) which is an absolute constant, but this includes all the terms (13)

$$\sum_{x=0}^{n} (p(x \mid \theta)) = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

and thus is of no value. It follows that whatever the sets of acceptance $X(\theta)$

the corresponding sum (14) will have values varying with the value of θ and hence the solution of the type (a) in the case of the binomial does not exist.

This, I think, gives the solution of the question raised by Professor Fisher. It is clear also that whenever the solution of the type (a) exists, it may be found by a suitable choice of sets of acceptance, and thus by the method explained in my earlier paper.

I should like now to raise another question. Past experience shows that the general problem of estimation may be formulated in different ways. The form of this problem as it appears in Bayes theorem, required for its solution the knowledge of the probabilities *a priori*.

The form of the same problem treated by R. A. Fisher in his theory of estimation was solved in terms of a new conception, that of likelihood.

The problem of estimation in its form of confidence intervals stands entirely within the bounds of the theory of probability, without involving any conception not already inherent in this theory. In the case of continuous distribution the problem also allows the solution (a) entirely independent of the probabilities *a priori*. Now it is shown that the necessity of the solution (b) is bound up with the nature of the problem if the distributions are discontinuous.

My question is: is it possible to formulate the problem of estimation in a fourth form, leading to a solution which (1) stands entirely on the grounds of the classical theory of probability, and (2) is not depending upon the probabilities *a priori*—whatever the conditions of the problem?

ANALYSIS OF VARIANCE CONSIDERED AS AN APPLICATION OF SIMPLE ERROR THEORY

By Walter A. Hendricks

The need for an elementary presentation of the methods of analysis of variance has been recognized by many investigators in various fields of research. A recent monograph by Snedecor (1934) is undoubtedly the most comprehensive attempt to satisfy this need which has appeared in the literature relating to the subject. Snedecor's treatment of the subject consists largely of the presentation of a number of standard types of problems to which the methods of analysis of variance are applicable, directions for performing the necessary computations, and a discussion of the conclusions which may be drawn from the data on the basis of the analysis.

In the opinion of the author of this paper, an elementary presentation of some of the theoretical considerations upon which the methods of analysis of variance are based would also be of some value. The methods of analysis of variance, as given by Fisher (1932), are presented as a natural consequence of intraclass correlation theory. However, the essential concepts may be presented in a more comprehensible form by the use of simple error theory.

It seems appropriate to begin such a presentation with a definition of variance. If we have an infinite number of measurements of the same quantity, the variance of a single measurement is defined as the arithmetic mean of the squares of the errors of those measurements. In actual practice, an infinite number of measurements can never be obtained. We have instead a sample of n measurements, $x_1, x_2, \cdots x_n$, from which the variance of a single measurement may be estimated. By referring to any text on the method of least squares, it may be verified that the best estimate, S^2, of the variance of a single measurement which can be obtained from a sample of n measurements is given by the equation:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

in which m represents the arithmetic mean of the n measurements. The quantity, $n - 1$, in the terminology of analysis of variance, is designated as the number of degrees of freedom available for estimating S^2.

It is often necessary to estimate S^2 from a number of different samples of measurements. In such cases, the best estimate of S^2 is obtained by calculating the weighted mean of the variances estimated from the individual samples, each variance being weighted by the number of degrees of freedom which were avail-

able for its estimation. The number of degrees of freedom upon which such an estimate of S^2 is based is given by the sum of these weights. Such an estimate of the variance of a single measurement is often designated as the variance "within samples."

In one of the simpler applications of analysis of variance, a number of samples of measurements are available, and the investigator is required to determine whether the magnitude of the quantity measured varied from sample to sample or whether all of the measurements may be regarded as having been made upon a quantity of the same magnitude.

An estimate, S^2, of the variance within samples may be obtained. Since S^2 is an estimate of the variance of a single measurement, the variance, S_i^2, of the arithmetic mean, m_i, of the measurements in any one sample is given by the equation:

$$S_i^2 = \frac{S^2}{n_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

in which n_i represents the number of measurements in the sample. Let there be r samples. Then another estimate, $S_i'^2$, of the variance of the mean, m_i, may be obtained from the observed distribution of the means, $m_1, m_2, \cdots m_r$, by the use of the formula for calculating the variance of a weighted observation as given in texts on the method of least squares:

$$S_i'^2 = \frac{1}{n_i(r-1)}[n_1(m_1 - m)^2 + n_2(m_2 - m)^2 + \cdots + n_r(m_r - m)^2] \ldots . (3)$$

in which:

$$m = \frac{n_1 m_1 + n_2 m_2 + \cdots + n_r m_r}{n_1 + n_2 + \cdots + n_r} \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Equations (2) and (3) yield two estimates of the variance of the mean, m_i. It is apparent that these two estimates will be equal, within the limits of sampling fluctuations, if all of the measurements in the r samples were made upon a quantity of the same magnitude. If the magnitude of the quantity measured varied from sample to sample, $S_i'^2$ will be greater than S_i^2. However, in actual practice, the two estimates of the variance of a particular mean are not compared directly. An equivalent comparison is made between two estimates of the variance of a single measurement. The first of these is nothing more than the variance within samples discussed earlier in this paper. The second estimate, which may be designated by S'^2, is the value which would have to be substituted for S^2 in equation (2) in order to make S_i^2 equal to the value given for $S_i'^2$ by equation (3). It is quite apparent that S'^2 may be found by the use of the equation:

$$S'^2 = \frac{1}{r-1}[n_1(m_1 - m)^2 + n_2(m_2 - m)^2 + \cdots + n_r(m_r - m)^2]. \ldots . (5)$$

S'^2 is often designated as the variance "between samples." A comparison of S'^2 with S^2 is obviously equivalent to a comparison of $S_i'^2$ with S_i^2.

If S'^2 is greater than S^2, a statistic, z, may be calculated:

$$z = \frac{1}{2} \log_e \frac{S'^2}{S^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

This statistic serves as a useful comparison between S'^2 and S^2 since its sampling distribution is known if all of the measurements comprising the data under investigation were made upon a quantity of the same magnitude. The distribution of z, under these conditions, is given by an equation of the form:

$$df = \frac{ke^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1+n_2)}} dz \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

in which n_1 represents the number of degrees of freedom available for estimating S'^2, and n_2 represents the number of degrees of freedom available for estimating S^2. It is apparent from equation (5) that $r - 1$ degrees of freedom are available for the estimation of S'^2 in the particular problem under discussion.

When any estimate of the variance of a single measurement is multiplied by the number of degrees of freedom available for making that estimate, the resulting product is known as a "sum of squares." The additive property of the sums of squares and the degrees of freedom contributes much to the elegance of the scheme of analysis just presented and is of considerable practical importance in problems of a type to be discussed later in this paper. In the case of the problem discussed above, the additive property of the sums of squares provides that the sum of the "sum of squares between samples" and the "sum of squares within samples" is equal to the sum of the squares of the deviations of all of the measurements from their arithmetic mean. The additive property of the degrees of freedom provides that the sum of the "degrees of freedom between samples" and the "degrees of freedom within samples" is equal to the "total degrees of freedom" which is nothing more than the total number of measurements diminished by unity.

The methods of analysis presented above may be applied to any study of the effects of a number of experimental treatments of the same kind upon the magnitude of a measurable quantity. If experimental treatments of more than one kind are imposed simultaneously, the effects of each may be studied by modifications of those methods. The discussion of those modifications, about to be presented in this paper, is limited to data which may be classified in an "$r \times s$" table, i.e., to studies of the effects of only two kinds of experimental treatments. More complex problems may be treated by simple extensions of the methods presented.

Consider an "$r \times s$" table composed of rs cells, each of which contains a number of measurements of some quantity. The magnitude of the quantity measured may vary from cell to cell, but the essential conditions under which the measurements were made must be the same for all cells. It is also under-

stood that no cell may be empty. Table 1 is an example of such a table. The individual measurements have not been represented. Only the number of measurements, n_{ij}, in each cell and the arithmetic mean, m_{ij}, of those measurements have been indicated. The arguments, a_i, represent r experimental treatments of one kind, and the arguments, b_j, represent s experimental treatments of another kind. The problem to be solved is to ascertain whether or not the differences among the experimental treatments of each kind had any effect on the magnitude of the quantity measured.

TABLE 1

Example of an "$r \times s$" Table Showing Only the Number of Measurements in Each Cell and the Arithmetic Mean of Those Measurements

	b_1	b_2	b_3	b_4		b_s
a_1	m_{11} n_{11}	m_{12} n_{12}	m_{13} n_{13}	m_{14} n_{14}		m_{1s} n_{1s}
a_2	m_{21} n_{21}	m_{22} n_{22}	m_{23} n_{23}	m_{24} n_{24}		m_{2s} n_{2s}
a_3	m_{31} n_{31}	m_{32} n_{32}	m_{33} n_{33}	m_{34} n_{34}		m_{3s} n_{3s}
a_r	m_{r1} n_{r1}	m_{r2} n_{r2}	m_{r3} n_{r3}	m_{r4} n_{r4}		m_{rs} n_{rs}

If each cell contains the same number of measurements, the effects of the experimental treatments indicated by the arguments, a_i, may be studied by comparing the variance "between rows" with the variance "within cells." The variance between rows may be calculated by regarding the r rows as r samples of measurements and applying an equation of the same form as equation (5). The variance within cells may be obtained by calculating the variance of a single measurement from the data in each cell separately and taking the mean of the resulting values. The effects of the experimental treatments indicated by the arguments, b_j, may be studied by comparing the variance "between columns" with the variance "within cells."

If the degrees of freedom between rows, between columns, and within cells are added, the sum will be less than the total number of degrees of freedom in the table. If the corresponding sums of squares are added, the sum is likely to be less than the total sum of squares. The differences are due to what is customarily designated as "interaction between rows and columns." The

more descriptive term, "differential response," is sometimes used to designate the same factor. The nature of this factor may be investigated by considering the effects of the experimental treatments, b_j, in each row of Table 1.

The data in each cell of Table 1 may be regarded as a sample of measurements. Therefore, the data in any row may be regarded as a set of s samples of measurements. By applying an equation of the same form as equation (5) to the data in any row, an estimate of the variance of a single measurement is obtained from the observed distribution of the means of the cells in that row. By calculating the arithmetic mean of the estimates for the r rows, an estimate of the variance of a single measurement is obtained from $r(s - 1)$ degrees of freedom. This estimate may be designated as the variance "between cells in the same row."

The variance between cells in the same row measures the average effect of differences among the experimental treatments, b_j, in individual rows. The variance between columns, which was discussed earlier in this paper, is calculated from $s - 1$ degrees of freedom and measures the effect of differences among the treatments, b_j, on the assumption that the effect of any one treatment upon the magnitude of the quantity measured was constant for every row. The number of degrees of freedom assignable to differential response of the various rows to the treatments, b_j, is $r(s - 1) - (s - 1)$ or $(r - 1)(s - 1)$. The sum of squares due to differential response is given by the difference between the sum of squares between cells in the same row and the sum of squares between columns. These relations follow from the additive property of degrees of freedom and sums of squares.

It may be observed that precisely the same results would be obtained by considering the effects of the treatments, a_i, in the various columns of Table 1. The degrees of freedom and sum of squares due to differential response of the various columns to the treatments, a_i, would be exactly equal to the corresponding values obtained for the differential response of the various rows to the treatments, b_j.

Up to this point the discussion has been concerned only with the special case in which each cell of Table 1 contains the same number of measurements. As a matter of fact, the methods given for the analysis of such data will yield correct results when applied to any "$r \times s$" table in which the numbers of measurements in the cells in every row are proportional to the corresponding marginal totals for the columns, and the numbers of measurements in the cells in every column are proportional to the corresponding marginal totals for the rows.

When the numbers of measurements in the various cells do not satisfy the above condition of proportionality, the distributions of the means of the rows and columns may be distorted, and, consequently, the methods of analysis described above may yield incorrect results. Efficient methods of analyzing such data have been presented by Yates (1933). A comprehensive discussion of these methods is considerably beyond the scope of this paper. One method,

described very briefly by Yates (1933) and designated as the "method of weighted squares of means," appealed to the author as being particularly valuable for practical work. No detailed discussion of the method seems to be available in the literature. Therefore, the following presentation may be of some interest.

Consider the experimental treatments represented by the arguments, a_i, in Table 1. It is necessary to find an average value for the magnitude of the quantity measured for each row of Table 1. However, this average must be of such a type that its value will not be distorted by the unequal numbers of measurements in the various cells. The unweighted arithmetic mean of the means of the cells in the row seems to be the logical average to use since, within the limits of sampling fluctuations, the value of this average will be identical with the value which would have been obtained if each cell had contained the same number of measurements. The averages for the r rows are:

$$m_{a_1} = \frac{1}{s}(m_{11} + m_{12} + \cdots + m_{1s})$$

$$m_{a_2} = \frac{1}{s}(m_{21} + m_{22} + \cdots + m_{2s})$$

$$\vdots$$

$$m_{a_r} = \frac{1}{s}(m_{r1} + m_{r2} + \cdots + m_{rs}). \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

By the law of propagation of error, the variance of any one of these unweighted means is given by the equation:

$$S^2_{\bar{a}_i} = \frac{1}{s^2}(S^2_{i1} + S^2_{i2} + \cdots + S^2_{is}) \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

in which $S^2_{\bar{a}_i}$ is the variance of m_{a_i}, and $S^2_{i1}, S^2_{i2}, \cdots, S^2_{is}$ are the variances of $m_{i1}, m_{i2}, \cdots, m_{is}$, respectively. If S^2 represents the variance of a single measurement, equation (9) may be written in the form:

$$S^2_{\bar{a}_i} = \left(\frac{1}{n_{i1}} + \frac{1}{n_{i2}} + \cdots + \frac{1}{n_{is}}\right)\frac{S^2}{s^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

The value of S^2 may be estimated from the individual measurements in the various cells. S^2 is nothing more than the variance within cells, as customarily calculated, and may be estimated from the $N - rs$ degrees of freedom within cells, in which N represents the total number of measurements in Table 1.

The variance of a single measurement may also be estimated from the observed distribution of the means of the type, m_{a_i}. These means are not of equal weight. Therefore, in order to find the variance of any one of them, it is first necessary to calculate the weighted mean of the r individual means. Since the weight of an arithmetic mean is inversely proportional to its variance, it is evident from

an inspection of equation (10) that the weight, p_{a_i}, of a mean, m_{a_i}, may be found from the equation:

$$\frac{1}{p_{a_i}} = \frac{1}{n_{i1}} + \frac{1}{n_{i2}} + \cdots + \frac{1}{n_{is}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

The weighted mean, m_a, may then be found:

$$m_a = \frac{p_{a_1} m_{a_1} + p_{a_2} m_{a_2} + \cdots + p_{a_r} m_{a_r}}{p_{a_1} + p_{a_2} + \cdots + p_{a_r}} \quad \ldots\ldots\ldots\ldots\ldots (12)$$

The variance $S'^2_{a_i}$, of any mean, m_{a_i}, as estimated from the observed distribution of means of this type, is given by:

$$S'^2_{a_i} = \frac{1}{p_{a_i}(r-1)} [p_{a_1}(m_{a_1} - m_a)^2 + p_{a_2}(m_{a_2} - m_a)^2 + \cdots + p_{a_r}(m_{a_r} - m_a)^2]. \quad \ldots\ldots (13)$$

By substituting $S'^2_{a_i}$ for $S^2_{a_i}$, and S^2_a for S^2, in equation (10) and solving the resulting equation for S^2_a, an estimate, S^2_a, of the variance of a single measurement is obtained from the observed distribution of means of the type, m_{a_i}. It is evident that, after making the indicated substitutions, equation (10) reduces to the form:

$$S^2_a = \frac{s^2}{r-1} [p_{a_1}(m_{a_1} - m_a)^2 + p_{a_2}(m_{a_2} - m_a)^2 + \cdots + p_{a_r}(m_{a_r} - m_a)^2] \ldots . (14)$$

It is interesting to observe that, if the numbers of measurements in the respective cells were equal, equation (14) would reduce to the formula for calculating the variance "between rows" as customarily applied in analysis of variance.

The two estimates, S^2 and S^2_a, of the variance of a single measurement may be compared in the usual manner by taking one-half of the natural logarithm of the ratio of the larger estimate to the smaller and making use of the tables of the values of "z" given by Fisher (1932). When using these tables, it is important to remember that S^2_a was estimated from $r - 1$ degrees of freedom.

The method of analysis just described may be employed to study the effects of differences among the experimental treatments indicated by the arguments, b_j, on the magnitude of the quantity measured. The unweighted means for the s columns are:

$$\begin{aligned} m_{b_1} &= \frac{1}{r}(m_{11} + m_{21} + \cdots + m_{r1}) \\ m_{b_2} &= \frac{1}{r}(m_{12} + m_{22} + \cdots + m_{r2}) \\ &\vdots \qquad\qquad \vdots \\ m_{b_s} &= \frac{1}{r}(m_{1s} + m_{2s} + \cdots + m_{rs}) \end{aligned} \quad \ldots\ldots\ldots\ldots\ldots (15)$$

The weight, p_{b_j}, of a mean of the type, m_{b_j}, may be found from the relation:

$$\frac{1}{p_{b_j}} = \frac{1}{n_{1j}} + \frac{1}{n_{2j}} + \cdots + \frac{1}{n_{rj}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (16)$$

A weighted mean, m_b, may be calculated:

$$m_b = \frac{p_{b_1}m_{b_1} + p_{b_2}m_{b_2} + \cdots + p_{b_s}m_{b_s}}{p_{b_1} + p_{b_2} + \cdots + p_{b_s}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (17)$$

An estimate, S_b^2, of the variance of a single measurement may be obtained from the observed distribution of means of the type, m_{b_j}, by the use of the equation:

$$S_b^2 = \frac{r^2}{s-1}\left[p_{b_1}(m_{b_1} - m_b)^2 + p_{b_2}(m_{b_2} - m_b)^2 + \cdots + p_{b_s}(m_{b_s} - m_b)^2\right] \ldots\ldots (18)$$

S_b^2 may be compared with S^2 in the usual manner.

If it is necessary to study the "interaction between rows and columns," the effects of the experimental treatments, b_j, may be studied for each individual row of Table 1. Consider the distribution of the means of the cells in a row designated by the argument, a_i. The weight of any one of these means is equal to the number of measurements in the cell. A weighted mean, m'_{a_i}, of the s means of cells in the row may be calculated:

$$m'_{a_i} = \frac{n_{i1}m_{i1} + n_{i2}m_{i2} + \cdots + n_{is}m_{is}}{n_{i1} + n_{i2} + \cdots + n_{is}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (19)$$

The variance, S'^2_{ij}, of the mean, m_{ij}, for any cell in the given row, as estimated from the observed distribution of means of this type, may be obtained from the equation:

$$S'^2_{ij} = \frac{1}{n_{ij}(s-1)}\left[n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2\right] \ldots\ldots (20)$$

The variance, S^2_{ij}, of the same mean, as estimated from the distribution of the individual measurements in the cell, may be obtained from the equation:

$$S^2_{ij} = \frac{S^2}{n_{ij}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

By substituting S'^2_{ij} for S^2_{ij}, and $S^2_{a_ib}$ for S^2, in equation (21) and solving the resulting equation for $S^2_{a_ib}$, an estimate, $S^2_{a_ib}$, of the variance of a single measurement is obtained from the observed distribution of the means of the cells in the given row. After making the indicated substitutions, equation (21) reduces to the form:

$$S^2_{a_ib} = \frac{1}{s-1}\left[n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2\right] \ldots\ldots (22)$$

Such an estimate, $S^2_{a_i b}$, of the variance of a single measurement may be obtained for each of the r rows in Table 1. By calculating the average, $S^2_{a\,b}$, of the variances of the type, $S^2_{a_i b}$, an estimate, $S^2_{a\,b}$, of the variance of a single measurement may be obtained from the $r(s-1)$ degrees of freedom between cells in the same row:

$$S^2_{a\,b} = \frac{1}{r(s-1)} \sum_{i=1}^{r} [n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2] \quad \ldots\ldots(23)$$

Equation (23) is identical with the formula for calculating the variance between cells in the same row as ordinarily applied in analysis of variance. This result is a direct consequence of the fact that the unequal numbers of measurements in the various cells had no distorting effect on the arithmetic means for individual cells.

The presence or absence of interaction may be verified by comparing $S^2_{a\,b}$ with S^2_b. In general, the actual variance due to interaction can not be obtained by the "weighted squares of means" method, for the various sums of squares do not possess the additive property when the analysis is made in this way. However, the comparison suggested above will yield sufficient information for most practical purposes.

For the special case in which r or s is equal to 2, the actual variance due to interaction may be calculated. Suppose $r = 2$ in Table 1. The following method, suggested by Yates (1933), yields an estimate of the variance due to interaction from a consideration of the differences, d_j, between the means of the two cells in each column:

$$\begin{aligned} d_1 &= m_{11} - m_{21} \\ d_2 &= m_{12} - m_{22} \\ &\vdots \qquad\quad \vdots \\ d_s &= m_{1s} - m_{2s} \end{aligned} \quad \ldots\ldots(24)$$

The variance, $S^2_{d_j}$, of any difference, d_j, is given by the equation:

$$S^2_{d_j} = \left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right) S^2 \quad \ldots\ldots(25)$$

The weight, p_j, of the difference, d_j, is given by the equation:

$$\frac{1}{p_j} = \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \quad \ldots\ldots(26)$$

The variance of the difference, d_j, as estimated from the observed distribution of differences, is given by the equation:

$$S'^2_{d_j} = \frac{1}{p_j(s-1)} [p_1(d_1 - d)^2 + p_2(d_2 - d)^2 + \cdots + p_s(d_s - d)^2] \quad \ldots(27)$$

in which:

$$d = \frac{p_1 d_1 + p_2 d_2 + \cdots + p_s d_s}{p_1 + p_2 + \cdots + p_s} \quad \ldots\ldots\ldots\ldots\ldots\ldots (28)$$

By means of these relations, an estimate, S_d^2, of the variance of a single measurement may be obtained from the observed distribution of the differences of the type, d_j. This estimate represents the variance due to interaction and may be obtained from the equation:

$$S_d^2 = \frac{1}{s-1}[p_1(d_1 - d)^2 + p_2(d_2 - d)^2 + \cdots + p_s(d_s - d)^2] \ldots\ldots (29)$$

It is quite apparent that $s - 1$ degrees of freedom are available for the estimation of the variance due to interaction in this particular example.

REFERENCES

FISHER, R. A., 1932. Statistical Methods for Research Workers, 4th edition. Edinburgh and London: Oliver and Boyd.

SNEDECOR, GEORGE W., 1934. Calculation and Interpretation of Analysis of Variance and Covariance. Ames, Iowa: Collegiate Press.

YATES, F., 1933. The principles of orthogonality and confounding in replicated experiments. Jour. Agr. Sci., 23: 108–145.

BUREAU OF ANIMAL INDUSTRY,
U. S. DEPARTMENT OF AGRICULTURE,
WASHINGTON, D. C.

NOTE ON THE DISTRIBUTIONS OF THE STANDARD DEVIATIONS AND SECOND MOMENTS OF SAMPLES FROM A GRAM-CHARLIER POPULATION

By G. A. Baker

T. N. Thiele in his "Theory of Observations" makes the following statement with regard to the distributions of the higher half-invariants in samples of n: "Not even for μ_2 have I discovered the general law of errors."[1] The purpose of this paper is to shed some light on the distribution of μ_2 and to give the distribution of second moments about a fixed point when the sampled population can be represented by a Gram-Charlier series.

The distribution of the second moments about a fixed point of samples is given in complete generality. It is known that if the sampled population is normal there is a simple relation between the distribution of the standard deviations of samples of n and the distribution of the second moments of the samples about the mean of the population. It was thought that such a relation might exist in case the sampled population could be represented by a Gram-Charlier series. Such is not the case. Again, it was thought that by obtaining the distribution of the standard deviations for samples of 2, 3, 4, $\cdots$ it might be possible to deduce empirically a general law of distribution. This proved an unfruitful line of investigation but required so much labor that the results should be reported to save others time and energy.

First, suppose that a population may be represented as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_5\varphi_5(x) + \cdots \tag{1}$$

where

$$\varphi_i(x) = \frac{d^i(e^{-\frac{1}{2}x^2})}{dx^i}.$$

Then applying Theorem II of the author's paper on "Random Sampling from Non-Homogeneous Populations"[2] we deduce at once the following theorem.

Theorem I. The distribution of the second moments about the origin of (1) of samples of n drawn at random from a population represented by (1) is precisely the same as the distribution of the second moments about the same point of samples of n drawn from a population represented by the first term of (1), that is a normal population, and is proportional to $x^{\frac{n-2}{2}} e^{-\frac{1}{2}x}$ (loc. cit.)

---

[1] Thiele, T. N., "The Theory of Observations," reprinted in the *Annals of Mathematical Statistics*, Vol. 2, No., 2, May, 1931, p. 208.

[2] *Metron*, Vol. 8, No. 3, Feb. 28, 1930.

This is not so surprising as it may seem at first if it is remembered that the odd subscript terms of a Gram-Charlier series slice off frequencies on one side of the mean of $a_0\varphi_0(x)$ and add them onto the other side in the same manner.

If we suppose that a population is given as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) + \cdots \tag{2}$$

in the same manner we get the following theorem.

THEOREM II. The distribution of the second moments measured from the origin of (2) of samples of n drawn at random from (2) will be a combination of distributions of the type of Theorem I with only even subscript terms contributing anything. The variations in the component distributions will consist of differences in the constant factors and the exponent of x, the estimate of the second moment. The lowest exponent will be $\frac{n-2}{2}$.

For instance, if

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{3}$$

and $n = 2$, the estimates of the second moment will be distributed as proportional to

$$e^{-\frac{1}{2}x}\left[(a_0+3)^2 - 12a_4(a_0+3)x + (36a_4^2 + 6a_0a_4 + 18a_4)\frac{x^2}{2!} - 36a_4^2\frac{x^3}{3!} + 9a_4^2\frac{x^4}{4!}\right].$$

Thus, it can be said that we know the distribution of the second moments of samples about a fixed point if the sampled population is of the Gram-Charlier type in the sense that given the number of terms necessary for an adequate representation and the number in the samples we can write down the desired distribution. However, this is not a simple matter. Further, if some relation existed between the distributions of the second moments about a fixed point and the standard deviations of the samples we would know the latter distribution also. Such a relation is not apparent for samples of 2 and 3.

Let us investigate the correlation surfaces of the means and standard deviations of samples of 2 and 3 drawn at random from a population represented by the first few terms of a Gram-Charlier series after the method of Dr. A. T. Craig.[3] The distributions of the standard deviations can then be obtained immediately by integration.

Suppose that

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{4}$$

[3] *Annals of Mathematical Statistics*, Vol. 3, No. 2, May, 1932, pp. 126–140.

and that we are considering samples of 2. The probability of the concurrence of x_1 and x_2 is

$$f(x_1)f(x_2) \tag{5}$$

and

$$\begin{aligned} x_1 &= -s + x \\ x_2 &= s + x \end{aligned} \tag{6}$$

where s is the standard deviation and x is the mean of a sample of 2. By means of (6), (5) becomes

$$\begin{aligned} (7)\qquad & e^{-(s^2+x^2)}[a_0^2 + a_0a_3(-6s^2x - 2x^3 + 6x) \\ & + a_0a_4(2s^4 + 12s^2x^2 - 12s^2 - 12x^2 + 6) \\ & + a_3^2(-s^6 + 3s^4x^2 + 6s^4 - 3s^2x^4 - 9s^2 + 9x^2 - 6x^4 + x^6) \\ & + a_3a_4(2s^6 - 6s^4x^3 - 6s^4x + 6s^2x^5 - 12s^2x^3 + 18s^2x - 2x^7 \\ & \qquad + 18x^5 - 42x^3 + 18x) \\ & + a_4^2(s^8 - 4s^6x^2 - 12s^6 + 6s^4x^4 + 12s^4x^2 + 42s^4 - 4s^2x^6 \\ & \qquad + 12s^2x^4 - 36s^2x^2 - 36s^2 + x^8 - 12x^6 + 42x^4 - 36x^2 + 9)]. \end{aligned}$$

To find the distribution of s we must integrate from $-\infty$ to ∞ with respect to x. Thus, (8) is obtained.

$$\begin{aligned} (8)\qquad & \sqrt{\pi}\, e^{-s^2}\left[a_0^2 + a_0a_4(2s^4 - 6s^2) + a_3^2\left(-s^6 + \frac{15}{2}s^4 - \frac{45}{4}s^2 + \frac{15}{8}\right)\right. \\ & \left. \qquad + 2a_3a_4s^6 + a_4^2\left(s^8 - 14s^6 + \frac{105}{2}s^4 - \frac{105}{2}s^2 + \frac{105}{2}\right)\right]. \end{aligned}$$

If we retain only two terms of (3), i.e. use

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) \tag{9}$$

and consider samples of 3 we obtain as the correlation surface of x and s

$$\begin{aligned} (10)\qquad & \frac{18\pi}{\sqrt{3}}\, se^{-\frac{1}{2}(3x^2+3s^2)}\left[a_0^3 - \frac{a_0^2a_3}{4}(-40x^3 + 24xs^2 - 24x)\right. \\ & \qquad + \frac{a_0a_3^2}{64}(-84s^6 + 525x^2s^4 - 2752x^4s^2 \\ & \qquad\qquad + 576s^4 - 1008x^2s^2 - 288s^2 - 5586x^6 + 270x^4 - 1728x^2) \\ & \qquad + \frac{a_3^3}{64}(28s^6 - 6189x^2s^4 - 28x^4s^2 - 629x^6 + 288s^4 + 1344x^4 \\ & \qquad\qquad \left. + 4608x^2s^2 - 288s^2 + 729x^2)\right]. \end{aligned}$$

The distribution of s can be obtained as before. The processes involved in obtaining (7) and (10) are so complicated that the general rule for writing the distribution of s is not apparent. Also, the relation of the distributions of s to the corresponding distributions of the second moments about a fixed point is not apparent.

In summary, the general distributions of the second moments about a fixed point of samples from a population represented by a definite number of terms of a Gram-Charlier series and the distributions of the standard deviations of samples of 2 and 3 from the same type of population are given and compared. No apparent relation exists between them.

ON THE FINITE DIFFERENCES OF A POLYNOMIAL

By I. H. Barkey

In this paper an apparently new and convenient method of finding the successive finite differences of a polynomial is considered. If operationally •

$$\phi(u + r_1 r_2) = E^{r_1 r_2}\,\phi(u) = (1 + \Delta r_1)^{r_2}\,\phi(u)$$

then for any polynomial $f(x)$ of degree "n"

$$\begin{aligned} f(x) &= p_0 x^n + p_1 x^{n-1} + \cdots + p_n \\ &= p_0(x + a)^n + q_{11}(x + a)^{n-1} + \cdots + q_{1n} \\ E^a f(x) &= p_0(x + a)^n + p_1(x + a)^{n-1} + \cdots + p_n \\ \Delta_a f(x) &= (p_1 - q_{11})(x + a)^{n-1} + (p_2 - q_{12})(x + a)^{n-2} + \cdots + (p_n - q_{1n}). \end{aligned}$$

Similarly, if $f_1(x) = \Delta a f(x)$, then

$$\begin{aligned} f_1(x) &= (p_1 - q_{11})(x + 2a)^{n-1} + q_{22}(x + 2a)^{n-2} + \cdots + q_{2n} \\ E^a f_1(x) &= (p_1 - q_{11})(x + 2a)^{n-1} + (p_2 - q_{12})(x + 2a)^{n-2} + \cdots + (p_n - q_{1n}) \\ \Delta_a f_1(x) &= (p_2 - q_{12} - q_{22})(x + 2a)^{n-2} + \cdots + (p_n - q_{1n} - q_{2n}) \end{aligned}$$

and so on for the higher orders, since $\Delta_a f_{s-1}(x) = \Delta_a^s f(x)$. In the practical application of this method, "a" may be conveniently taken as unity, and an abridged form of synthetic division employed. Thus, if

$f(x) = 5x^4 + 3x^3 + 7x^2 - 2x + 3$, then

5 +	3 +	7 −	2 +	3 = f
−	2 +	9 −	**11** +	14
−	7 +	16 −	27	
−	12 +	28		
−	17			
	20 −	21 +	25 −	**11** = f_1
		− 41 +	**66** −	77
		− 61 +	127	
		− 81		
		60 −	102 +	66 = f_2
			− **162** +	228
			− 222	
			120 −	**162** = f_3
				− 282
				120 = f_4.

As is evident from the darkened numerals, all figures to the right of the dotted line are redundant and may be omitted. From the above,

$$\Delta f(x) = 20(x+1)^3 - 21(x+1)^2 + 25(x+1) - 11$$

$$\Delta^2 f(x) = 60(x+2)^2 - 102(x+2) + 66$$

$$\Delta^3 f(x) = 120(x+3) - 162$$

$$\Delta^4 f(x) = 120.$$

SOME PRACTICAL INTERPOLATION FORMULAS

By John L. Roberts

Sometimes we wish to find by means of interpolation an approximation to a particular value of w_x in the interval between the known values, w_0 and w_1. But it also might be desirable in the interval from w_0 to w_1 to interpolate several approximations to w_x at equidistant values of x. It is very important to know that a formula which might be very satisfactory to interpolate a particular value in an interval might seriously fail to be the most satisfactory formula when it is desired to interpolate several values in the same interval. The range of this paper is so limited that we only wish to find by means of interpolation several approximations to the true value of w_x in the interval from w_0 to w_1 at equidistant values of x.

One way to perform an interpolation of this sort is to use osculatory interpolation.[1] The real function of osculatory interpolation is to secure smoothness at the known points, which are sometimes called pivotal points. By roughness is meant that one or more of the successive derivatives are discontinuous at the pivotal points. Experience proves that the osculatory formulas usually secure smoothness either at the expense of labor or by a loss of accuracies over the entire range from w_0 to w_1. Frequently the function of interpolation formulas is to save labor. In many cases it appears reasonable to save labor by a loss of both smoothness and accuracy. Formulas are herein selected, without direct regard for smoothness, so as to secure the best possible compromise between a maximum of accuracy and a minimum of labor. It appears that this results in many cases in a loss of smoothness that is no more objectionable than the loss in accuracy.

The actuarial profession, while trying to perfect their methods of constructing mortality tables, have made contributions of a high order of scholarship to the theory of osculatory interpolation. But since the statistician, the astronomer, the physicist, and other scientists also have occasions to make interpolations, it seems to be very important to discuss the problem of finding the most practical methods of interpolation, not only from the special viewpoint of the actuary, but also from the general viewpoint of mathematics.

Δw_x is called the first difference of w_x, and may be defined by $\Delta w_x = w_{x+1} - w_x$.

[1] Since this paper presupposes certain knowledge on the part of the reader, it may be worth while to indicate some sources of this knowledge. The elementary parts of this knowledge can be found in any good book on finite differences. "Population Statistics and Their Compilation" by Hugh H. Wolfenden, published by the Actuarial Society of America, contains an excellent summary of osculatory interpolation. This summary indicates some valuable sources of information.

Second, third, and higher differences are merely successive differences of the first. When use is made of central difference interpolation formulas, it is convenient to adopt Woolhouse's notation, which is defined by means of the following equations: $\Delta w_{-2} = a_{-2}$, $\Delta w_{-1} = a_{-1}$, $\Delta w_0 = a_1$, $\Delta w_1 = a_2$, $\Delta^2 w_{-2} = b_{-1}$, $\Delta^2 w_{-1} = b_0$, $\Delta^2 w_0 = b_1$, $\Delta^3 w_{-2} = c_{-1}$, $\Delta^3 w_{-1} = c_1$, $\Delta^4 w_{-2} = d_0$, $\Delta^5 w_{-2} = e_1$, $\Delta^6 w_{-3} = f_0$, etc.

An important family of curves can be represented by

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)B + \frac{1}{6}x(x-1)\left(x-\frac{1}{2}\right)C. \tag{1}$$

Assume $u_0 = w_0$ and $\Delta u_0 = \Delta w_0$. Then a study of (1) shows that a_1, which has already been defined, must be a factor in the second term in order that (1) may be satisfied when $x = 1$. (1) is a third degree equation. However, if $C = 0$, (1) becomes a second degree equation; if both $B = 0$ and $C = 0$, (1) becomes a first degree equation. In other words, by giving B and C proper values, (1) can be made to become many different interpolation formulas.

For many purposes interpolation by a first degree formula is not sufficiently accurate. We, therefore, might wish to interpolate by either a second or a third degree formula. Since it is possible to draw an unlimited number of second degree curves or third degree curves between the points P_0 and P_1, the problem of selecting the best second degree interpolation curve and the best third degree curve is of great practical importance.

I

Suppose that w_{-2}, w_{-1}, w_0, w_1, w_2, and w_3 can be found in a table of values of the function w_x, and that we wish to find by means of interpolation several approximate values of w_x in the interval from w_0 to w_1. These six given values of w_x can be used to determine six pivotal points, which determine a fifth degree curve. Suppose this curve represents the function v_x. Then w_x and v_x would have exactly the same values at the six pivotal points, but would have values which are only approximately the same at other points. Using the first six terms of the Gauss central difference interpolation formula, we have

$$\begin{aligned} v_x = v_0 + xa_1 + \frac{1}{2!}x(x-1)b_0 + \frac{1}{3!}(x+1)x(x-1)c_1 \\ + \frac{1}{4!}(x+1)x(x-1)(x-2)d_0 \\ + \frac{1}{5!}(x+2)(x+1)x(x-1)(x-2)e_1. \end{aligned}$$

It is proper to use in this formula the differences a_1, b_0, etc., which have already been defined as differences of w_x because these differences are exactly equal to the corresponding differences of v_x. Suppose P_0, $P_{\frac{1}{2}}$, $P_{\frac{1}{2}}$, and P_1 are four points

which are determined by v_x. Then B and C can be determined so that (1) will represent the curve which can go through these four points.
Then

$$u_{\frac{1}{3}} = u_0 + \frac{1}{3}a_1 - \frac{1}{9}\left(B - \frac{1}{18}C\right)$$

and

$$v_{\frac{1}{3}} = u_0 + \frac{1}{3}a_1 - \frac{1}{9}\left(b_0 + \frac{4}{9}c_1 - \frac{5}{27}d_0 - \frac{7}{81}e_1\right).$$

Also

$$u_{\frac{2}{3}} = u_0 + \frac{2}{3}a_1 - \frac{1}{9}\left(B + \frac{1}{18}C\right)$$

and

$$v_{\frac{2}{3}} = u_0 + \frac{2}{3}a_1 - \frac{1}{9}\left(b_0 + \frac{5}{9}c_1 - \frac{5}{27}d_0 - \frac{8}{81}e_1\right).$$

Since $u_{\frac{1}{3}} = v_{\frac{1}{3}}$ and $u_{\frac{2}{3}} = v_{\frac{2}{3}}$, we have two equations, which can be solved for B and C.

$$B = b - \frac{5}{27}d \text{ and } C = c_1 - \frac{1}{9}e_1 \tag{2}$$

where b and d are defined by

$$b = \frac{1}{2}(b_0 + b_1) \text{ and } d = \frac{1}{2}(d_0 + d_1).$$

A study of (1) shows that $u_{\frac{1}{3}}$ does not depend upon C because the term containing C becomes zero when $x = \frac{1}{2}$, and also shows that u_x over the entire range from u_0 to u_1 is more sensitive to errors in B than errors in C. The B in (2) usually contains some error because the six terms of the Gauss formula which were used in determining B usually produce results which are only approximate. Consequently a comparatively large error in C would not produce an important error.
Assume

$$B = b - \frac{5}{27}d \text{ and } C = c_1 - \frac{5}{27}e_1. \tag{3}$$

B is the same in both (2) and (3), but C is not the same. The accuracy of (2) and the accuracy of (3) do not differ by an important amount. On the other hand, if any attempt to apply (2) is compared with the working illustrations of (3) in this article, it will be found that (2) to an important extent is more laborious than (3). Therefore (3) is a better compromise between a maximum of accuracy and a minimum of labor than (2). For this reason (2)

ought not to be regarded as a practical formula. On the other hand (2) because of its great accuracy serves as an ideal with which other formulas can be compared. In other words (2) is of theoretical importance.

In like manner another interpolation formula can be found if we use the first four terms of the Gauss formula to determine $P_{\frac{1}{2}}$. Then

$$u_{\frac{1}{2}} = u_0 + \frac{1}{2}a_1 - \frac{1}{8}B$$

and

$$v_{\frac{1}{2}} = u_0 + \frac{1}{2}a_1 - \frac{1}{8}\left(b_0 + \frac{1}{2}c_1\right).$$

Since $u_{\frac{1}{2}} = v_{\frac{1}{2}}$, we can solve for B, and C is left arbitrary. If $C = 0$, we again get an excellent compromise between a maximum of accuracy and a minimum of labor. The following second degree formula results.

$$B = b \text{ and } C = 0. \tag{4}$$

In order that the value of (3) and (4) may be appreciated, they are herein compared with some other formulas which have been of historical importance.

If the point $P_{\frac{1}{2}}$ can first be accurately determined, a second degree curve through the points P_0, $P_{\frac{1}{2}}$, and P_1 would probably give more accurate results than such a curve through the points P_0, P_1, and P_2 because the first three points are in a smaller neighborhood; the second curve can be represented by the first three terms of the Gregory-Newton interpolation formula. The points P_{-1}, P_0, P_1, and P_2 determine a third degree curve, which can be represented by the first four terms of the Gauss central difference formula. It is probable that these terms would determine $P_{\frac{1}{2}}$ much more accurately than the first three terms of the Gregory-Newton formula because the latter is not a central difference formula with respect to $P_{\frac{1}{2}}$ and because four terms usually give more accurate results than only three terms. Consequently there is a strong probability that (4) is more accurate than the first three terms of the Gregory-Newton formula. In like manner (4) is more accurate than the first three terms of the Gauss formula. It is interesting to observe that (4) is the first three terms of the Newton-Bessel formula.

$$\text{If } B = b \text{ and } C = 3c_1,$$

then (1) is equivalent to Karup's osculatory interpolation formula in terms of differences taken centrally. B is the same in both (4) and Karup's formula. No interpolation formula can be very accurate unless C is about equal to c_1. Since, then, the error in C in Karup's formula is about twice as great as the error in C in (4), his formula is distinctly less accurate than (4). Since (4) is a second degree curve and Karup's formula is a third degree curve, his formula is very much more laborious. (4) is extremely accurate for a formula having its labor saving properties; for many purposes its roughness and inaccuracy appear to

be in about the right proportion. On the other hand Karup's formula is extremely inaccurate for a formula so laborious; its only good point is its smoothness.

Changing somewhat the meanings of u and w, (3) may be written

$$u_{x+n} = u_n + x\Delta u_n + \frac{1}{2}x(x-1)\left[\frac{1}{2}(\Delta^2 w_n + \Delta^2 w_{n-1}) - \frac{5}{54}(\Delta^4 w_{n-1} + \Delta^4 w_{n-2})\right] + \frac{1}{6}x(x-1)\left(x - \frac{1}{2}\right)\left(\Delta^3 w_{n-1} - \frac{5}{27}\Delta^5 w_{n-2}\right).$$

If

$$\frac{du}{dx} = u'_{x+n},$$

then

$$u'_{0+0} - u'_{1-1} = \frac{1}{54}d_0 - \frac{5}{162}f_0,$$

which is the amount of discontinuity in $\frac{du}{dx}$ at P_0. (3) has greater smoothness than (4); in other words (3) is more like an osculatory formula. On the other hand

$$B = b - \frac{1}{6}d \text{ and } C = c_1 - \frac{1}{6}e_1, \tag{5}$$

which is equivalent to an important osculatory interpolation formula by Mr. Robert Henderson, compares much better with (3) from the viewpoint of labor saving and accuracy than Karup's formula does with (4).

II

An excellent formula can be easily spoiled if the method of applying it is not practical. Mr. Henderson, in the Transactions of the Actuarial Society of America, Vol. IX, applies (5) in such a way that the numerical work is very convenient. Some writers seem to have been very careless about this matter. A method intended to interpolate several values between w_0 and w_1 should provide that the end value w_1 shall be exactly reproduced if no error is made in the computation. In other words a good method should provide a check upon the work. At the same time, in order to avoid unnecessary labor, the work should not retain unnecessary decimal places or figures. In other words fictitious accuracy should be avoided. The following working illustrations are intended to show good methods of application of formulas and to show how much labor is necessary in order to apply them; also the size of the errors can be used to illustrate the theory.

When (4) is applied at either end of the table, where terms are not available for the calculation of the differences required, it should be assumed that the fourth differences that cannot be computed vanish and the required differences should be filled in consistently with that assumption. Δw_x represents the first differences. But it is convenient to have S represent the first differences in such a manner that they are arranged centrally in the working illustration. S^2 in like manner represents the second differences. The 2 in S^2 means S^2 is a second difference, and does not have the familiar meaning used in algebra. In the case of (4), $\Delta u_x = a_1 + xB$, $\Delta^2 u_x = B$, and the higher differences all equal zero. Since we wish in the working illustration of (4) to interpolate four values between w_0 and w_1, δ and δ^2 are defined by $\delta u_x = u_{x+.2} - u_x$ and $\delta^2 u_x = \delta u_{x+.2} - \delta u_x$. It is proved in any good book on finite differences that there are possibilities that Δ and δ, which are symbols of operation, can be separated from the functions upon which they operate, and they can be treated as if they were algebraic numbers. Consequently $1 + \delta = (1 + \Delta)^{\frac{1}{2}}$. In other words by means of the binomial law $\delta u_x = (.2\Delta - .08\Delta^2)u_x$, where all the terms within the parenthesis are to be considered as operating upon u_x. Also $\delta^2 u_x = .04\Delta^2 u_x$. s, s_x, and s^2 are defined by $s = s_x = \delta u_x$, and $s^2 = s^2_{.4} = \delta^2 u_x$. Therefore the middle $s = \delta u_{.4} = .2a_1$, and $s^2 = .04B = .02(b_0 + b_1)$. We are now in position to apply (4) to the case when $w_x = (1.04)^n$. It might prevent confusion if it is stated that x and n are related to each other in such a way that we always interpolate between w_0 and w_1.

n	$(1.04)^n$	s	S	S^2	s^2
80	23.050	.9218		.845	
81	23.9718	.9603			
82	24.9321	.9988	4.994		.0385
83	25.9309	1.0373			
84	26.9682	1.0758			
85	28.044	1.1190		1.081	
86	29.1630	1.1670			
87	30.3300	1.2150	6.075		.0480
88	31.5450	1.2630			
89	32.8080	1.3110			
90	34.119	1.3636		1.317	
91	35.4826	1.4210			
92	36.9036	1.4784	7.392		.0574
93	38.3820	1.5358			
94	39.9178	1.5932			
95	41.511			1.553	

Some of the explanation of the application of (4) applies to (3) and does not need to be repeated. The method herein used of applying (3) is either the same as or a development of the Henderson method of applying (5). If it is desired to apply (3) at either end of the table, where terms are not available for the calculation of the differences required, it can be assumed that the sixth differences that can not be computed vanish and the required differences can be filled in consistently with that assumption. A study of the theory underlying this assumption shows that it does not result in a true central difference formula and that it consequently results usually in some loss of accuracy. In the case of (3) before the finding of the differences of (1), it is convenient to write it as follows:

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)\left(B + \frac{1}{2}C\right) + \frac{1}{6}x(x-1)(x-2)C\,.$$

Then

$$\Delta u_x = a_1 + x\left(B + \frac{1}{2}C\right) + \frac{1}{2}x(x-1)C\,,$$

$$\Delta^2 u_x = \left(B + \frac{1}{2}C\right) + xC, \text{ and } \Delta^3 u_x = C\,.$$

Suppose we wish to interpolate four values between w_0 and w_1. δ and δ^2 have already been defined. $\delta^3 u_x = \delta^2 u_{x+.2} - \delta^2 u_x$. Then $1 + \delta = (1 + \Delta)^{\frac{1}{2}}$, or $\delta u_x = (.2\Delta - .08\Delta^2 + .048\Delta^3)u_x$. Also $\delta^2 u_x = (.04\Delta^2 - .032\Delta^3)u_x$ and $\delta^3 u_x = .008\Delta^3 \cdot$ s^2, s^2_x, and s^3 are defined by $s^2 = s^2_x = \delta^2 u_{x-.2}$, and $s^3 = s^3_{.4} = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.2} = .04\left(B - \frac{1}{2}C\right) = .04\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.8} = .04\left(B + \frac{1}{2}C\right) = .04\left(b_1 - \frac{5}{27}d_1\right).$$

.1852 might be a useful approximation to $\frac{5}{27}$. The remaining s^2, s should be filled in so that they are in arithmetical progression with irregularities at the ends. If the irregularities can be distributed equally at both ends, the irregularities cause an error in C, but none in B. Errors in B are more important than those in C. The middle $s = \delta u_{.4} = .2a_1 - s^3$. In the following working illustration, $w_x = \sin n$.

n	$\sin n$	S	S^2	S^3	S^4
−60	−.86603				
		.36603			
−30	−.50000		.13397		
		.50000		−.13397	
0	.00000		.00000		.00000
		.50000		−.13397	
30	.50000		−.13397		.03588
		.36603		−.09809	
60	.86603		−.23206		
		.13397			
90	1.00000				

n	$\sin n$	s	s^2	s^3
0	.00000	.104498	.000000	
6	.104498	.103374	−.001124	
12	.207872	.101125	2249	−.001125
18	.308997	.097751	3374	
24	.406748	93252	4499	
30	.50000		−.005624	

Suppose we wish to interpolate nine values between w_0 and w_1 by the use of (3). Then $\delta u_x = u_{x+.1} - u_x$, $\delta^2 u_x = \delta u_{x+.1} - \delta u_x$, and $\delta^3 u_x = \delta^2 u_{x+.1} - \delta^2 u_x$. Consequently $1 + \delta = (1 + \Delta)^{1/10}$, or $\delta u_x = (.1\Delta - .045\Delta^2 + .0285\Delta^3)u_x$. Then $\delta^2 u_x = (.01\Delta^2 - .009\Delta^3)u_x$ and $\delta^3 u_x = .001\Delta^3$. $s^2 = s_x^2 = \delta^2 u_{x-.1}$ and $s^3 = s_{.4}^3 = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.1} = .01\left(B - \frac{1}{2}C\right) = .01\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.9} = .01\left(B + \frac{1}{2}C\right) = .01\left(b_1 - \frac{5}{27}d_1\right).$$

$\delta u_{.4} = (.1a_1 - 4s^3) - \frac{1}{2}\delta^2 u_{.4}$ and $\delta u_{.5} = (.1a_1 - 4s^3) + \frac{1}{2}\delta^2 u_{.4}$.

n	sin n	s	s^2	s^3
0	.00000	52318	.000000	
3	.052318	52179	−.000139	
6	.104497	51899	280	
9	.156396	51478	421	
12	.207874	.050916	562	−.000141
15	.258790	.050212	703	
18	.309002	49368	844	
21	.358370	48383	985	
24	.406753	47257	1126	
27	.454010	45990	1267	
30	.50000		−.001406	

Suppose we wish to interpolate five values between w_0 and w_1. The first $s^2 = \frac{1}{36}\left(b_0 - \frac{5}{27}d_0\right)$ and the last $s^2 = \frac{1}{36}\left(b_1 - \frac{5}{27}d_1\right)$.

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta u_x) - \frac{1}{2}\delta^2 u_{\frac{1}{2}}$$

and

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta^3 u_x) + \frac{1}{2}\delta^2 u_{\frac{1}{2}}.$$

In the following working illustration the given values of sin n are written correct to five decimal places; in other words after each decimal point there are five symbols or digits representing numbers; also each of these symbols is written in the scale of ten. It can be observed that some values of u_x, s, s^2, and s^3 in the working illustration have six symbols to the right of the decimal point, and that some values have seven symbols to the right of the decimal point. In all cases the sixth symbol to the right of the decimal point is written in the scale of ten, and the seventh symbol is written in the scale of six. This procedure provides a check by exactly reproducing w_1. Also this procedure does not cause much fictitious accuracy, and can be quickly used after a little practice.

n	sin n	s	s^2	s^3
0	.00000	87130	.000000	
5	.0871305	86479	−.000651	
10	.1736104	.0851775	1302	
15	.2587883	.0832245	1953	−.000651
20	.3420132	80620	2604	
25	.4226341	77365	3255	
30	.50000		−.003906	

In general if we wish to interpolate $i - 1$ values between w_0 and w_1 when i is neither five nor ten, w_1 can be exactly reproduced if some of the symbols are written in the scale of i. If $i = 12$, it is evident that we need two extra symbols, say t and e, to stand for ten and eleven respectively. If we wish to interpolate $i - 1$ values between w_0 and w_1 by the use of (4), in the computation each of u_x, s and s^2 except the given values should contain one more symbol than each given value contains, and the extra symbol should be written in the scale of i.

ON EVALUATING A COEFFICIENT OF PARTIAL CORRELATION

By Grace Strecker

It is to be shown here that when the multiple correlation coefficient $R_{n;12\cdots(n-1)}$ is found by the method of Horst[1] the partial correlation coefficient $R_{n(n-1);12\cdots(n-2)}$ can be found in terms of the β's. If we are interested only in evaluating a partial correlation between two variables, we may also employ the method which will be given here.

Without loss of generality the dependent variables may be chosen to be the nth and $(n-1)$st. The coefficient of partial correlation as given by Rietz[2] may be expressed in the following form:

$$(1) \qquad R_{n(n-1);12\cdots(n-2)} = \sqrt{\frac{\dfrac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} - \dfrac{R}{R_{nn}}}{\dfrac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}}}}.$$

$R_{(n-1)(n-1)}$ may be treated as a new determinant R'. Regarding its elements as the coefficients of a set of normal equations ($n-1$ in all) whose constant terms are zero, we may follow through the Doolittle elimination process. For the case where $n = 4$ we have the table given below.

In comparing this outline with the one illustrating the Doolittle elimination process for R when $n = 4$ we see that

$$\gamma'_{11} = \gamma_{11} = \frac{A_{11}}{R^2},$$

$$\gamma'_{22} = \gamma_{22} = \frac{rA_{1122}}{R^2 A_{11}},$$

$$\gamma'_{33} = \alpha'_{33} - \sum_{2}^{3} \beta'_{i3} = \alpha_{44} - \sum_{2}^{3} \beta_{i4}.$$

Therefore, we have

$$R' = \frac{A_{11}}{R^2} \cdot \frac{rA_{1122}}{R^2 A_{11}} \cdot \left(\alpha_{44} - \sum_{2}^{3} \beta_{i4}\right)$$

$$= \prod_{1}^{2} \gamma_{ii} \left(\alpha_{44} - \sum_{2}^{3} \beta_{i4}\right).$$

---

[1] Horst Paul, *A Short Method for Solving for a Coefficient of Multiple Correlation*, Annals of Mathematical Statistics, Vol. III, No. 1, Feb. 1932, pp. 40–44.

[2] Rietz, H. L., *Mathematical Statistics*, p. 101.

Reciprocal	1	2	3	α	β	γ	δ
$-\frac{R^2}{A_{11}}$	$\frac{A_{11}}{R^2}$	$\frac{A_{12}}{R^2}$	$\frac{A_{14}}{R^2}$	α_1'		γ_1'	
	-1	$-\frac{A_{12}}{A_{11}}$	$-\frac{A_{14}}{A_{11}}$				δ_1'
		$\frac{A_{22}}{R^2}$	$\frac{A_{24}}{R^2}$	α_2'			
		$-\frac{A_{12}^2}{R^2A_{11}}$	$-\frac{A_{12}A_{14}}{R^2A_{11}}$		β_{22}'		
$-\frac{R^2A_{11}}{AA_{1122}}$		$\frac{AA_{1122}}{R^2A_{11}}$	$\frac{AA_{1124}}{R^2A_{11}}$			γ_2'	
		-1	$-\frac{A_{1124}}{A_{1122}}$				δ_2'
			$\frac{A_{44}}{R^2}$	α_3'			
			$-\frac{A_{14}^2}{R^2A_{11}}$		β_{23}'		
			$-\frac{AA_{1124}^2}{R^2A_{11}A_{1122}}$		β_{33}'		
			$\alpha_3' - \sum_2^3 \beta_{i3}'$			γ_3'	
			-1				δ_3'

In the general case:

$$\gamma_{11}' = \gamma_{11},$$

$$\gamma_{22}' = \gamma_{22},$$

$$\vdots$$

$$\gamma_{(n-2)(n-2)}' = \gamma_{(n-2)(n-2)},$$

$$\gamma_{(n-1)(n-1)}' = \alpha_{nn} - \sum_2^{n-1} \beta_{in}.$$

Hence

$$R_{(n-1)(n-1)} = R' = \prod_{1}^{n-2} \gamma_{ii} \left(\alpha_{nn} - \sum_{2}^{n-1} \beta_{in} \right).$$

Since $R = \prod_{1}^{n} \gamma_{ii}$, then $R_{(n-1)(n-1)nn} = \prod_{1}^{n-2} \gamma_{ii}$, from which we see that

$$\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} = \frac{\prod_{1}^{n-2} \gamma_{ii} \left(\alpha_{nn} - \sum_{2}^{n-1} \beta_{in} \right)}{\prod_{1}^{n-2} \gamma_{ii}} = \alpha_{nn} - \sum_{2}^{n-1} \beta_{in}.$$

But since $\alpha_{nn} = 1$, then

$$\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} = 1 - \sum_{2}^{n-1} \beta_{in}.$$

It has been shown that

$$\frac{R}{R_{nn}} = 1 - \sum_{2}^{n} \beta_{in}.$$

Substituting the above values for $\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}}$ and $\frac{R}{R_{nn}}$ in equation (1), we have

$$R_{n(n-1);12\cdots(n-2)} = \sqrt{\frac{1 - \sum_{2}^{n-1} \beta_{in} - \left(1 - \sum_{2}^{n} \beta_{in} \right)}{1 - \sum_{2}^{n-1} \beta_{in}}},$$

or

$$R_{n(n-1);12\cdots(n-2)} = \sqrt{\frac{\beta_{nn}}{1 - \sum_{2}^{n-1} \beta_{in}}}.$$

Hence it is seen that when the β's given by Horst (page 42) are calculated, it is an easy matter to solve for the partial correlation $R_{n(n-1);12\cdots(n-2)}$.

ST. LOUIS UNIVERSITY.

A THEORY OF VALIDATION FOR DERIVATIVE SPECIFICATIONS AND CHECK LISTS[1]

By Lee Byrne

Visiting Professor of Secondary Education, New York University

Part I. Research Products Which May Be Classified as Derivative Specifications and Check Lists

Meaning of Specification

In specification something is assigned a specific character. The something to be thus assigned a specific character may be called the specificandum. The specific character assigned to the specificandum, or (as a second meaning) the act of so doing, may be called the specification.

A proposition is the smallest unit in which it is possible to embody a complete thought and is ordinarily represented by a single sentence. In specification the characterization may be confined to a single proposition or it may be extended to include an indefinitely large number of propositions. So a specification may be embodied in a sentence, a paragraph, a chapter, or a whole book. No matter how far it is extended it will never give complete determination, as our knowledge cannot be made exhaustive or our control be given an absolute precision.

In view of the meaning assigned to specification it is evident that very many books and monographs could in this sense be classified as specifications.

Meaning of Derivative Specification

There is a type of specification (book or monograph) which is developed by deriving it from a group or class of specifications which already exist. This class may be a total class of all such specifications, or a group of those accepted as authoritative, or a group of those taken to be representative. A specification derived in this manner may be called a derivative specification. As an example we could take almost any first-class work by a present-day historian; by historians it would be called "secondary" because it is based on study of pre-existent documents called "primary sources."

Meaning of Check List

The act of *deriving* a product from a pre-existent set of documents may, as we have seen, take the form of a derivative specification, embracing an as-

[1] This paper is an amplification of a report made in the statistical section of the American Educational Research Association at its meeting in February, 1931.

semblage of determinates or determinations. On the other hand the product derived may be intended merely to indicate the ground covered or to be covered by determination, without actually selecting the particular determinations. Such a product will be called a check list. The term is not a very happy one, but it is in very common use. If we think of a specification as an assemblage of determinations then a check list could be thought of as a corresponding set of determinables.[2] Since any determinable is capable of an indefinite number of determinations it is evident that a long check list could give rise to an extremely large number of different specifications, of which, of course, some fraction might prove undesirable, inadmissible, or false.

Modes of Specification: How We Specify

If we examine any specification to see how the specifying is done we shall find that it ultimately takes the form of specification under aspects. The following diagram indicates the principal (perhaps all the) possibilities in the way of specification.

Naming the original or main specificandum
  Naming an aspect
    Characterization of the specificandum under the aspect named
    .
    .
    .
  Naming a relation (includes process, operation etc.)
    Naming an aspect of the relation
      Characterization of the relation under aspect named
      .
      .
      .
    Naming a relatum or thing related (a new specificandum)
      Naming an aspect of the relatum
        Characterization of relatum under aspect named
      .
      .
      .
  Naming a part
    Naming an aspect of the part
      Characterization of the part under aspect named
  .
  .
  .

(The naming of aspects may be merely implicit but it is always present in principle.)

---

[2] On the notion of the "determinable," which is due to W. E. Johnson, see his Logic, Cambridge University Press (1921), Part I, p. xxxv and Chapter XI.

Thus it appears that if specification is pressed far enough it always ultimately becomes specification under aspects. Aspect and determinable may be regarded as synonyms.

Current Examples of Derivative Specifications and Check Lists

At the present time it will be found that we have very many products of research which take forms capable of being classified as some kind of derivative specification or (derivative) check list in the senses in which these expressions have been explained.

I have distinguished more than twenty different logical types of derivative specification or check list which are exemplified in the current literature of educational research and related subjects. However space will not permit exhibition of examples of these different types.

Part II. Validation of Derivative Specifications and Check Lists

Many research products may be classified as derivative specifications or check lists, derivative in the sense that they have been derived from a group of documents (books, articles, journals, newspapers, courses of study, etc.) through analysis of their content. Such source documents themselves we shall call specifications or groups of specifications.

The only validation problem raised here is the question whether the resulting check list or derivative specification truly represents the class of source specifications used. The further question whether the class of source specifications itself constitutes a satisfactory source is not discussed.

From this point of view, if a check list or derivative specification is based in some suitable manner on *all* the documents of the class represented, no real validation problem arises; the validity has to be regarded as perfect.

It may often happen that the investigator does not wish to analyse *all* of the specifications of the class in question but prefers to save time and labor by confining his analysis to a select group drawn from the total class as a sample. In this case the problem arises as to how far results based on such sample should be judged to be truly representative of the entire class of specifications (most of which have not been analysed). A problem of this nature may be called the problem of validity for this kind of work.

Such a validation problem appears to take the same form whether the product to be validated is a derivative specification or (derivative) check list. Accordingly we shall for the sake of brevity carry on the discussion by referring to the problem as that of validating (derivative) check lists. The same principles would apply if the product happened to be a derivative specification.

In order to consider the validity of a check list based on a sample group of specifications (called here a Sample Check List) we may hypothesize a check list based in the same manner on the entire class of specifications from which the sample was drawn. Such a hypothetical check list (which is not made) will be called the Ideal Check List. Then the problem of validity may be con-

ceived as the question as to how far the content of the Sample Check List agrees with the unknown content of the Ideal Check List.

An overlapping of the two appears ordinarily to be certain but a failure of complete coincidence is very highly probable. The question is what degree of coincidence is to be expected.

This general validity problem naturally divides into two separate questions. The first question asks what proportion of the content of the Sample Check List may be expected to be present also in the Ideal Check List; this may be called the (sub-) problem of reliability. The second question asks what proportion of the content of the Ideal Check List may be expected to be present in the Sample Check List; this may be called the (sub-) problem of completeness. The answers to these two problems, if expressed in numerical percentages, could be called the Index of Reliability and Index of Completeness respectively.

We shall first consider these two problems in their simplest form and afterward in a more complex form in which they exhibited themselves in a recent study by the writer.[3] The simple case presents no great difficulty and it is possible that a different method of disposing of it might be preferred. The more complex case, however, appears to be rather difficult of solution and the writer has not been able to find in the literature any developed technique for handling it. The simple case is presented here primarily because it affords, by further extension, a successful approach to the difficult problem of the more complex case.

Simple Case

Terms and Symbols

The "class of specifications" will be understood to consist of all specifications which belong to the whole class of specifications regarded as a source, a class which we claim to represent in our final product. In this problem the "class" will not be regarded as indefinitely large but as consisting of a definite number of specifications, a number to be ascertained by actual count or by careful estimate.

"Sample specifications" are the limited group selected from the class for purposes of actual analysis, and which play the rôle of representing the whole class. The remaining specifications of the class are not analyzed.

"Sample Check List Material" is a name for the assemblage of all the different items found in one or more sample specifications.

"Ideal Check List Material" is a name for a hypothetical assemblage of all the different items found in one or more specifications in the class. Only those appearing in some sample specifications can be actually known, the rest are hypothetical.

---

[3] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

Write

M (constant) = total number of specifications in class
N (variable) = number of these specifications in which a particular item under consideration appears (this number is hypothetical and some of the particular items themselves are hypothetical)
m (constant) = number of sample specifications
n (variable) = number of sample specifications in which a particular (the same) item appears

Values of n may be expected to vary for different items, from m to 0 by intervals of 1, the zero value appertaining to any item wholly absent from the Sample Check List Material (hypothetically present in Ideal Check List Material).

Values of N might be expected to vary, for different items, from M to 1 by intervals of 1. But in this problem the convention will be adopted that the range is from M downward by intervals of $\frac{M}{m}$. Thus if the number M should be five times as large as the number m then the range for N would be treated as proceeding from M downward by intervals of 5: $M, M - 5, M - 10, \cdots 5$.

A "tabulation" will mean a statistical table showing how many different items appear in every possible number of specifications. A tabulation must be made by actual count for the items of the sample specifications, and will show the number of items having each possible value of n. A similar tabulation is hypothetical for the items in all the specifications of the class, that is for the number of items having each value of N permitted by the convention of the last paragraph.

"Tabulation cell" (or simply "cell") will mean, as needed, either the number of items or the group of items appearing in any designated number of specifications. For Sample Check List Material it will be the number or group of items to which a particular value of n appertains; for Ideal Check List similarly the number of items or group of items to which a particular value of N appertains (hypothetically).

"Sample Check List" will mean a list of items selected from the Sample Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be, for example, the numerical ratio $n \geqq \frac{m}{2}$.

"Ideal Check List" will mean a list of items selected from the Ideal Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be the numerical ratio $N \geqq \frac{M}{2}$.

Problem of Reliability

The problem of reliability may be restated and renamed the General Reliability Problem. This may be broken up into a group of problems which will

be called Elementary Reliability Problems. Each of the latter may be in turn broken up into a group of problems which will be called Ultimate Reliability Problems. Each Ultimate Reliability Problem may be solved directly. Combination of these solutions will yield solutions of the Elementary Reliability Problems. Combinations of the latter solutions will finally yield the solution of the General Reliability Problem.

These problems will now be stated

General Reliability Problem: What proportion of the items present in Sample Check List may be expected to be present also in Ideal Check List?

Elementary Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in Ideal Check List?

Ultimate Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in some designated cell in Ideal Check List?

To solve an Ultimate Problem:

From the Fundamental Theorem in the Theory of Inductive Probability (Whittaker, E. T. and Robinson, G. The Calculus of Observations. London: Blackie & Son. 1924. p. 305) the solution may be expressed as

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

Whittaker and Robinson's statement of the Fundamental Theorem in the Theory of Inductive Probability is as follows (form slightly changed without change in meaning):

"Suppose that a certain observed phenomenon may be accounted for by any one of a certain number of hypotheses, of which one, and not more than one, must be true: suppose moreover that the probability of the R-th hypothesis, as based on information in our possession before the phenomenon is observed, is P_R, while the probability of the observed phenomenon, on the assumption of the truth of the R-th hypothesis, is p_s. Then when the observation of the phenomenon is taken into consideration, the probability of the R-th hypothesis is

$$\frac{P_R \cdot p_s}{\Sigma Pp}$$

where the symbol Σ denotes the summation over all the hypotheses."[4]

It is clear that an Ultimate Reliability Problem is a case falling under this Fundamental Theorem. The observed phenomenon is any item occurring in any specified cell of Sample Check List, say cell $n = s$. It may be accounted for by a certain number of hypotheses as to its source in the Ideal Check List

[4] For the fundamental position of this theorem in a theory of science and for its proof one may also consult Jeffreys, H. Scientific Inference. Cambridge: Cambridge University Press. 1931. Chapter II (section 2.34).

Material; the different cells in the Ideal Check List Material are these different hypotheses of origin, hypothetical because we do not *know* from which one it has come but only that it must have come from some one of them; the cell from which it actually comes is the true hypothesis, though we do not know which one that is. That the origin of the item is in cell $N = R$ is the R-th hypothesis, and its probability is written P_R. The probability of the occurrence of the phenomenon on the assumption of the truth of the R-th hypothesis is the probability that an item in cell $N = R$ will appear in Sample Check List in cell $n = s$ and its probability is written p_s. As we clearly have in our Ultimate Reliability Problem a case falling under the Fundamental Theorem quoted we may accept as the required solution of the Ultimate Reliability Problem the formula already given in the initial statement:

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

This expresses the probability that any item found in Sample-Check-List cell $n = s$ comes from (and appears in) Ideal-Check-List-Material cell $N = R$, or it gives the proportion of items found in Sample-Check-List cell $n = s$ that may be expected to come from (or appear in) Ideal-Check-List-Material cell $N = R$.

Meaning of any value of P (say P_R) = the probability that any item, drawn at random from those cells of Ideal Check List Material which are possible sources of items in Sample-Check-List cell $n = s$, will happen to be drawn from cell $N = R$.

Meaning of any value of p (say p_s) = the probability that any item in Ideal-Check-List cell $N = R$ will also be present in Sample-Check-List cell $n = s$. (Important: this supposition is *not* equivalent to its converse.)

Evaluation of P_R:

$$P_R = \frac{\text{number of items in cell } N = R}{\text{number of items in all cells which are possible sources of items in cell } n = s}.$$

For this ratio it is necessary to assume that the shape of the numerical curve formed by the group of Ideal-Check-List-Material cells is the same as that of the numerical curve formed by the group of Sample-Check-List-Material cells. On this assumption we may replace the numerator by the number of items in the Sample-Check-List-Material cell having an abscissa corresponding to that of the Ideal-Check-List-Material cell $N = R$, and replace the denominator by the sum of the numbers of items in all the cells with abscissae corresponding to those of Ideal-Check-List-Material cells which are possible sources of items in cell $n = s$.

Evaluation of p_s:

By the aid of "the definition of probability which is used in practically all treatises on the subject" (Coolidge, J. L. An Introduction to Mathematical

Probability. Oxford: Oxford University Press. 1925. p. 4) and the principle underlying the Theory of Combinations (Whitworth, W. A. Choice and Chance. New York: G. E. Stechert & Co. 1927. Proposition II) we are able to arrive at the evaluation:

$$p = \frac{C_{m-n}^{M-N} C_n^N}{C_m^M}$$

in which, for any p (say p_s), we employ for N the value $N = R$, and for n the value $n = s$. As the denominator later cancels out it may be disregarded throughout, simplifying the formula to

$$p = C_{m-n}^{M-N} C_n^N .$$

(A symbol such as C_n^N is read "the number of combinations of N things taken n at a time"; also written in several other forms.)

The definition referred to may be worded as follows (Coolidge's own preferred definition is not quite the same):

"An event can happen in a certain number of ways, which are all equally likely. A certain proportion of these are classed as *favorable*. The ratio of the number of favorable ways to the total number is called the probability that the event will turn out favorably."

The principle underlying the Theory of Combinations may be quoted from Whitworth as follows (also found in ordinary works on algebra):

"If one operation can be performed in m ways, and then a second can be performed in n ways, and then a third in r ways, (and so on), the number of ways of performing all the operations will be $m \times n \times r \times$ etc."

If it is not at once clear that the formula for evaluation of p follows from the definition and principle just quoted, the following considerations should make it evident.

We are working in terms of a particular item belonging to a particular Ideal-Check-List-Material cell, say cell $N = R$. "Favorable" occurrence requires that this item fall in a particular Sample-Check-List cell, say $n = s$, while falling in any other Sample-Check-List-Material cell (including cell $n = 0$ for absence) is "unfavorable." Again the real meaning of the "favorable" occurrence is that the item will be found in just $n = s$ out of the m specifications of the sample, and absent in the remaining $m - n$ specifications of the sample. Moreover presence in Ideal-Check-List-Material cell $N = R$ means that the item occurs in just $N = R$ of the M specifications that constitute the whole class and is absent in $M - N$ of these specifications. The total number of all the ways (favorable and unfavorable) in which our event can happen means the same as the total number of all the ways in which a group of m specifications can be selected from a larger group of M, and this is, of course, written C_m^M and given us in our denominator. The number of favorable ways in which our event can happen means the same as the number of ways in which N specifications containing the item can form groups of n specifications while at the

same time $M - N$ specifications not containing the item can form groups of $m - n$ specifications; the first distribution can be done in C_n^N ways and the second in C_{m-n}^{M-N} ways, so by Whitworth's principle the number of ways which these things can happen simultaneously is $C_{m-n}^{M-N} C_n^N$. Assembling numerator and denominator we have the formula initially stated for evaluation of p, viz.:

$$p = \frac{C_{m-n}^{M-N} C_n^N}{C_m^M}.$$

This is the general formula; in applying to the particular example $N = R, n = s$ the replacements for N and n, of course, give

$$p_s = \frac{C_{m-s}^{M-R} C_s^R}{C_m^M}.$$

Having a means of evaluating P and p we may solve all needed Ultimate Problems. The resulting solutions of the needed Ultimate Reliability Problems (not necessarily completed) enables us to arrive at the solution of any needed Elementary Reliability Problem in the form of a percentage which may be called an Index of Reliability for the Sample-Check-List cell in question. In computing this percentage we distinguish source-cells that belong to the Ideal Check List from other source-cells that belong to the Ideal Check List Material but not to the Ideal Check List.

By properly averaging cell-Indices of Reliability (which are really Indices of Reliability for the individual items in the cells) we may obtain a solution of the General Problem of Reliability in the form of an Average Index of Reliability for the Sample Check List as a whole.

In addition to the Average Index of Reliability for *the* Sample Check List we may easily secure also Average Indices of Reliability for any series of briefer Sample Check Lists selected from *the* Sample Check List, by properly averaging the Indices of cells contained in any Sample Check List in question, keeping the original criterion for Ideal Check List.

In practice it may not be necessary to compute all cell-Indices, as a portion of these may be entered in tables by any methods of interpolation regarded as acceptable.

Problem of Completeness

Again we have General, Elementary, and Ultimate Problems. These may be stated as follows:

General Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in Sample Check List?

Elementary Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

Ultimate Completeness Problem: What proportion of the items in a particular cell in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

To solve an Ultimate Problem:

From principles already used the proportion to be expected is the same as the value of p alone in an Ultimate Reliability Problem, viz.:

$$\frac{C_{m-n}^{M-N} C_n^N}{C_m^M}.$$

By the use of this formula we may solve the Ultimate Problems for all values of N represented in Ideal Check List and all values of n represented in Sample Check List; some of these solutions will have a value of zero.

For each value of n, if we properly average the solutions of the Ultimate Problems, we obtain a solution of the Elementary Problem for one Sample-Check-List cell in the form of a percentage which may be called the Index of Completeness for the particular Sample-Check-List cell. In securing this average it is necessary to multiply each Ultimate Problem solution by a relative number corresponding to the assumed ratio of number of items in the particular Ideal-Check-List cell to the number of items in all the Ideal-Check-List cells. The source of the assumed relative numbers is the same as that used in evaluating P in the Reliability Problem.

When we have an Index of Completeness for each Sample-Check-List cell we may obtain a Total Index of Completeness for the Sample Check List as a whole by summing the cell-Indices of Completeness of all the cells of the Sample Check List. By an equivalent but preferable method we may divide the last-named result by the sum of the cell-Indices of Completeness of all the cells of the Sample Check List Material (including cell $n = 0$); by this method the C_m^M of the original formula cancels out and so may be disregarded throughout.

A Total Index of Completeness is similarly obtainable for *a* Sample Check List (any Sample Check List selected from *the* Sample Check List) by summing the cell-Indices of Completeness of the appropriate cells. Thus, if desired, a tabulation may be made showing Indices of Completeness for a series of Sample Check Lists differing in extent.

A combined tabulation may show for each of a series of Sample Check Lists its Index of Reliability and its Index of Completeness.

More Complex Case

So far we have considered a validation problem of simple type. In the writer's Check List Materials for Public School Building Specifications[5] a more complex problem was presented, due to the introduction of the concept of the Applicable Case. A Check List for School Building Specifications was developed with a view to its use by school officials or others as an aid in judging proposed school building specifications with reference to their completeness or incompleteness of determination. The position was taken that a new specification ought not to be charged with the omission of a given item unless the building (as repre-

[5] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

sented by the specification) had an Applicable Case for that item. To give a single example, the Check List contains various items relating to the specifying of marble work. It did not seem appropriate to score a specification down for the omission of numerous determinations in marble work, if in fact there was no marble in the building to be determined. This situation is expressed by saying that there are no Applicable Cases for those items.

It seems likely that there are other research problems in which the question ought to be raised whether adequate treatment does not require the introduction of the concept of the Applicable Case. If so a more difficult validation problem is presented than would otherwise be the case.

In the more complex case indicated solution is obtained by making the necessary extensions in the procedures followed for the simple case.

Modifications in Terms and Symbols

M (constant) = total number of specifications in class
D (variable) = number of these specifications containing an Applicable Case for a particular item
N (variable) = number of the latter specifications which also contain the particular item
m (constant) = number of specifications in sample
d (variable) = number of these specifications containing an Applicable Case for the particular item
n (variable) = number of the latter specifications which also contain the particular item

Values of d range from m to 0 by intervals of 1, and those of n range from d to 0 by intervals of 1.

The convention is adopted that values of D range from M downward, and those of N from D downward, by intervals of $\frac{M}{m}$.

(Tabulation) cell will mean the number of items (or the group of items) having a common value of d and a common value of n.

The criterion for membership in the Sample Check List may, for illustrative purposes, be taken as $n \geqq \frac{d}{2}$.

The criterion for membership in the Ideal Check List may, for illustrative purposes, be taken as $N \geqq \frac{D}{2}$.

Problem of Reliability

Following the same principle and line of reasoning as for the simple case we arrive at the same general formula for the solution of an Ultimate Reliability Problem, viz.:

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

Meanings of values of P and p are the same as before except that cells must be described respectively in terms of n and d values instead of n values alone, or N and D values instead of N values alone.

P_R is evaluated in the same manner as before, using the new meaning of "cell."

For p_s the evaluation now becomes

$$p = \frac{C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N}{C_m^M}$$

which through cancellation may be simplified to the working formula

$$p = C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N.$$

The reasoning leading to the denominator C_m^M is unchanged and so this denominator itself remains unchanged. The numerator for the evaluation of p is altered to the extent shown by the consideration that, in producing "favorable" ways, we now have to do with the number of simultaneous possibilities of drawing n specifications from a group of N specifications containing a particular item, drawing $d - n$ specifications from a group of $D - N$ specifications which contain an Applicable Case for this particular item but do not contain this item itself, and of drawing $m - d$ specifications from a group of $M - D$ specifications which contain no Applicable Case for the item.

Problem of Completeness

Following the same principles and line of reasoning as for the simple case we arrive at the following formula for the solution of an Ultimate Completeness Problem:

$$\frac{C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N}{C_m^M}.$$

By suitable treatment bringing about cancellations the working formula may be reduced to

$$C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N$$

Techniques and Aids in Computation

The present paper is limited to an attempt to explain with adequate fullness the proposed theory of validation for derivative specifications and check lists, and space is lacking in which to exhibit techniques of actual computation. One specimen problem worked out in fairly complete detail, together with remarks on available aids in computation will be found in Appendix A3 in typewritten copies of the writer's "Check List Materials for Public School Building Specifications" on file in the Library of Teachers College, Columbia University; the Appendices are not included in the printed edition.

A NOTE ON SHEPPARD'S CORRECTIONS

By Solomon Kullback

In this note we shall derive a simple relation between the characteristic function of the grouped distribution and the characteristic function of the original continuous distribution, assuming that the frequency curve has high contact with the x-axis at both ends.

If we set $p_s = \int_{x_s-\frac{w}{2}}^{x_s+\frac{w}{2}} f(x)\,dx$, then the characteristic function of the grouped distribution is given by

$$\psi(t) = \sum e^{itx_s} p_s \tag{1}$$

where $i = \sqrt{-1}$. Replacing p_s by its value as given above, we have

$$\begin{aligned}\psi(t) &= \sum e^{itx_s} \int_{x_s-\frac{w}{2}}^{x_s+\frac{w}{2}} f(x)\,dx \qquad (2)\\ &= \sum e^{itx_s} \int_{-\frac{w}{2}}^{\frac{w}{2}} f(x+x_s)\,dx\\ &= \int_{-\frac{w}{2}}^{\frac{w}{2}} dx \sum e^{itx_s} f(x+x_s)\\ &= \sum e^{itx_s} f(x_s) \int_{-\frac{w}{2}}^{\frac{w}{2}} e^{-itx}\,dx\,.\end{aligned}$$

There is no difficulty about justifying the inversion of the order of integration and summation.

Because of the assumption of high-contact with the axis of x at both ends of the frequency curve, we have

$$\varphi(t) = \int e^{itx} f(x)\,dx = w \sum e^{itx_s} f(x_s) \tag{3}$$

so that

$$\psi(t) = \frac{2}{wt} \sin \frac{tw}{2}\, \varphi(t)\,. \tag{4}$$

This is the desired result, from which there follows the desired moment relations by equating coefficients of $(it)^r$ on both sides of the equation. For example:

$$1 + M_1 it + \frac{M_2}{2!}(it)^2 + \frac{M_3}{3!}(it)^3 + \cdots = \left(1 + \frac{(it)^2 w^2}{4}\frac{1}{3!} + \frac{(it)^4 w^4}{16}\frac{1}{5!} + \cdots\right)$$

$$\left(1 + m_1 it + \frac{m_2}{2!}(it)^2 + \cdots\right)$$

$$= 1 + m_1 it + \frac{(it)^2}{2!}\left(m_2 + \frac{w^2}{12}\right) + \frac{(it)^3}{3!}\left(m_3 + \frac{m_1 w^2}{4}\right) + \cdots$$

or

$$M_1 = m_1\,; \qquad M_2 = m_2 + \frac{w^2}{12}\,; \qquad M_3 = m_3 + \frac{m_1 w^2}{4}\,; \cdots$$

WASHINGTON, D. C.

THE LIMITING DISTRIBUTIONS OF CERTAIN STATISTICS[1]

By J. L. Doob

There have been many advances in the theory of probability in recent years, especially relating to its mathematical basis. Unfortunately, there appears to be no source readily available to the ordinary American statistician which sketches these results and shows their application to statistics. It is the purpose of this paper to define the basic concepts and state the basic theorems of probability, and then, as an application, to find the limiting distributions for large samples of a large class of statistics. One of these statistics is the tetrad difference, which has been of much concern to psychologists.

I

Let $F(x)$ be a monotone non-decreasing function, continuous on the left, defined at every point of the x-axis, and satisfying the conditions

$$\lim_{x\to-\infty} F(x) = 0\,, \qquad \lim_{x\to\infty} F(x) = 1\,. \tag{1}$$

Then the function $F(x)$ is said to be the distribution function of a chance variable $\mathbf{x}$, and $F(x)$ is said to be the probability that $\mathbf{x} < x$. The curve $y = F(x)$ is sometimes called the ogive in statistics. The chance variable $\mathbf{x}$ itself is merely the function x, taken in conjunction with the monotone function $F(x)$.

If $\int_{-\infty}^{\infty} x\,dF(x)$ exists as an absolutely convergent Stieltjes integral, the value of the integral is called the expectation of $\mathbf{x}$, and will be denoted by $E(\mathbf{x})$.

II

Let $F(x_1, \cdots, x_n)$ be a function defined over n-dimensional space, which is monotone, non-decreasing, continuous on the left in each coördinate if the others are held fast, and which satisfies the conditions

$$\lim_{x_j\to-\infty} F(x_1, \cdots, x_n) = 0, \quad j = 1, \cdots, n, \qquad \lim_{x_1,\cdots,x_n\to\infty} F(x_1, \cdots, x_n) = 1 \tag{2}$$

where in the last limit, $x_1, \cdots, x_n$ become infinite together. Then $F(x_1, \cdots, x_n)$ is said to be the distribution function of a set of chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and $F(x_1, \cdots, x_n)$ is said to be the probability that all the inequalities $\mathbf{x}_j < x_j$, $(j = 1, \cdots, n)$, hold simultaneously. It can be shown that the function $F_j(x) = \lim_{\xi_1,\cdots,\xi_{n-1}\to\infty} (\xi_1, \cdots \xi_{j-1}, x, \xi_j, \cdots \xi_{n-1})$ is of the type discussed in §I. The

[1] Research under a grant-in-aid from the Carnegie Corporation.

function $F_j(x)$ is called the distribution function of $\mathbf{x}_j$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are called independent if $F(x_1, \cdots, x_n) = \prod_{j=1}^{n} F_j(x_j)$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are merely the functions $x_1, \cdots, x_n$ defined over n-dimensional space, taken in conjunction with the function $F(x_1, \cdots, x_n)$.

If $a_1, \cdots, a_n$ are any real numbers, the number $F(a_1, \cdots, a_n)$, the probability that $\mathbf{x}_j < a_j$, $j = 1, \cdots, n$, is also called the probability that a sample $(x_1, \cdots, x_n)$ shall be in the region of n-dimensional space determined by $x_j < a_j$, $j = 1, \cdots, n$. Thus regions of this special type have probabilities attached to them. Using the usual additivity rules, probabilities can be attached to more general regions, and in fact probability can be defined on a collection C of regions including all open sets, closed sets and all sets which can be obtained from them by repeatedly taking sums, products, and complements. (Such point sets are called Borel measurable). The resulting function of point sets is non-negative and completely additive.[2]

If $f(x_1, \cdots, x_n)$ is any function of $x_1, \cdots, x_n$ let E_x be the set of points $(x_1, \cdots, x_n)$ where $f < x$. Suppose that E_x is in the collection C for all values of x, and let $F(x)$ be the probability attached to the set E_x. Then it is readily seen that $F(x)$ has the properties discussed in §I and is therefore the distribution function of a new chance variable $\mathbf{x}$, which will be denoted by $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. The chance variable $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is merely the function $f(x_1, \cdots, x_n)$ taken in conjunction with the distribution function $F(x_1, \cdots, x_n)$. (An example is $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$, determining the chance variable $\mathbf{x}_1 + \cdots + \mathbf{x}_n$.) Suppose that $E(\mathbf{x})$ exists,

$$E(\mathbf{x}) = \int_{-\infty}^{\infty} x dF(x). \tag{3}$$

Then it can be shown that the n-dimensional (Lebesgue)-Stieltjes integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dF(x_1, \cdots, x_n) \tag{4}$$

exists and has the value $E(\mathbf{x})$. Conversely the existence of the integral (4) implies that of (3).

If there is a Lebesgue-integrable function $\varphi(x_1, \cdots, x_n)$ such that

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \varphi(x_1, \cdots, x_n)\, dx_1 \cdots dx_n, \tag{5}$$

---

[2] That is, if $p(E)$ is the value of the set function on the set E, and if $E_1, E_2, \cdots$ are point sets with no common points, and which are in C, $p\left(\sum_{m=1}^{\infty} E_m\right) = \sum_{m=1}^{\infty} p(E_m)$.

THE LIMITING DISTRIBUTIONS OF CERTAIN STATISTICS[1]

By J. L. Doob

There have been many advances in the theory of probability in recent years, especially relating to its mathematical basis. Unfortunately, there appears to be no source readily available to the ordinary American statistician which sketches these results and shows their application to statistics. It is the purpose of this paper to define the basic concepts and state the basic theorems of probability, and then, as an application, to find the limiting distributions for large samples of a large class of statistics. One of these statistics is the tetrad difference, which has been of much concern to psychologists.

I

Let $F(x)$ be a monotone non-decreasing function, continuous on the left, defined at every point of the x-axis, and satisfying the conditions

$$\lim_{x\to-\infty} F(x) = 0\,, \qquad \lim_{x\to\infty} F(x) = 1\,. \tag{1}$$

Then the function $F(x)$ is said to be the distribution function of a chance variable $\mathbf{x}$, and $F(x)$ is said to be the probability that $\mathbf{x} < x$. The curve $y = F(x)$ is sometimes called the ogive in statistics. The chance variable $\mathbf{x}$ itself is merely the function x, taken in conjunction with the monotone function $F(x)$.

If $\int_{-\infty}^{\infty} x\,dF(x)$ exists as an absolutely convergent Stieltjes integral, the value of the integral is called the expectation of $\mathbf{x}$, and will be denoted by $E(\mathbf{x})$.

II

Let $F(x_1, \cdots, x_n)$ be a function defined over n-dimensional space, which is monotone, non-decreasing, continuous on the left in each coördinate if the others are held fast, and which satisfies the conditions

$$\lim_{x_j\to-\infty} F(x_1, \cdots, x_n) = 0, \quad j = 1, \cdots, n, \qquad \lim_{x_1,\cdots,x_n\to\infty} F(x_1, \cdots, x_n) = 1 \tag{2}$$

where in the last limit, $x_1, \cdots, x_n$ become infinite together. Then $F(x_1, \cdots, x_n)$ is said to be the distribution function of a set of chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and $F(x_1, \cdots, x_n)$ is said to be the probability that all the inequalities $\mathbf{x}_j < x_j$, $(j = 1, \cdots, n)$, hold simultaneously. It can be shown that the function $F_j(x) = \lim_{\xi_1,\cdots,\xi_{n-1}\to\infty} (\xi_1, \cdots \xi_{j-1}, x, \xi_j, \cdots \xi_{n-1})$ is of the type discussed in §I. The

[1] Research under a grant-in-aid from the Carnegie Corporation.

function $F_j(x)$ is called the distribution function of $\mathbf{x}_j$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are called independent if $F(x_1, \cdots, x_n) = \prod_{j=1}^{n} F_j(x_j)$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are merely the functions $x_1, \cdots, x_n$ defined over n-dimensional space, taken in conjunction with the function $F(x_1, \cdots, x_n)$.

If $a_1, \cdots, a_n$ are any real numbers, the number $F(a_1, \cdots, a_n)$, the probability that $\mathbf{x}_j < a_j$, $j = 1, \cdots, n$, is also called the probability that a sample $(x_1, \cdots, x_n)$ shall be in the region of n-dimensional space determined by $x_j < a_j$, $j = 1, \cdots, n$. Thus regions of this special type have probabilities attached to them. Using the usual additivity rules, probabilities can be attached to more general regions, and in fact probability can be defined on a collection C of regions including all open sets, closed sets and all sets which can be obtained from them by repeatedly taking sums, products, and complements. (Such point sets are called Borel measurable). The resulting function of point sets is non-negative and completely additive.[2]

If $f(x_1, \cdots, x_n)$ is any function of $x_1, \cdots, x_n$ let E_x be the set of points $(x_1, \cdots, x_n)$ where $f < x$. Suppose that E_x is in the collection C for all values of x, and let $F(x)$ be the probability attached to the set E_x. Then it is readily seen that $F(x)$ has the properties discussed in §I and is therefore the distribution function of a new chance variable $\mathbf{x}$, which will be denoted by $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. The chance variable $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is merely the function $f(x_1, \cdots, x_n)$ taken in conjunction with the distribution function $F(x_1, \cdots, x_n)$. (An example is $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$, determining the chance variable $\mathbf{x}_1 + \cdots + \mathbf{x}_n$.) Suppose that $E(\mathbf{x})$ exists,

$$E(\mathbf{x}) = \int_{-\infty}^{\infty} x dF(x). \tag{3}$$

Then it can be shown that the n-dimensional (Lebesgue)-Stieltjes integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dF(x_1, \cdots, x_n) \tag{4}$$

exists and has the value $E(\mathbf{x})$. Conversely the existence of the integral (4) implies that of (3).

If there is a Lebesgue-integrable function $\varphi(x_1, \cdots, x_n)$ such that

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \varphi(x_1, \cdots, x_n)\, dx_1 \cdots dx_n, \tag{5}$$

---

[2] That is, if $p(E)$ is the value of the set function on the set E, and if $E_1, E_2, \cdots$ are point sets with no common points, and which are in C, $p\left(\sum_{m=1}^{\infty} E_m\right) = \sum_{m=1}^{\infty} p(E_m)$.

the function φ is said to be the density function of the distribution. In this case (4) becomes

$$(4') \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, \varphi(x_1, \cdots, x_n)\, dx_1 \cdots dx_n.$$

The probability attached to a point set E in the collection C is the integral (4) (or (4′) if there is a density function), where $f = 1$ over E and $f = 0$ elsewhere.

III

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables. We suppose that for every integer n, $\mathbf{x}, \mathbf{x}_n$ determine a bivariate distribution. Then it is readily seen from §II that there is a chance variable $|\, \mathbf{x}_n - \mathbf{x}\,|$ and therefore that $P\{|\, \mathbf{x}_n - \mathbf{x}\,| \leqq \lambda\}$[3] is defined for every number λ. If

$$(6) \qquad \lim_{n \to \infty} P\{|\, \mathbf{x}_n - \mathbf{x}\,| \leqq \lambda\} = 1$$

for every positive number λ, the sequence x_n is said to converge stochastically, or to converge in probability, to $\mathbf{x}$. If α is a constant, $P\{|\, \mathbf{x}_n - \alpha\,| \leqq \lambda\}$ is also defined for every number λ, and there is a corresponding definition of stochastic convergence to α. The usual theorems about limits hold: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge stochastically to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges stochastically to $\mathbf{x} + \mathbf{y}$, etc.

An example of stochastic convergence is given by the law of large numbers. Let $\mathbf{x}$ be a chance variable with distribution function $F(x)$ and suppose that $E(\mathbf{x})$, $E(\mathbf{x}^2)$ exist, i.e. that

$$\int_{-\infty}^{\infty} x dF(x), \qquad \int_{-\infty}^{\infty} x^2 dF(x)$$

are absolutely convergent integrals. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be chance variables whose n-variate distribution function is $\prod_{j=1}^{n} F(x_j)$: we are thus supposing that the variables all have the same distribution and form an independent set. Then $\frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$ is a new chance variable, and Tchebycheff's inequality furnishes an immediate proof that $\frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$ converges stochastically to $E(\mathbf{x})$.[4]

---

[3] Throughout this paper, if γ represents a set of conditions on chance variables, $P\{\gamma\}$ will denote the probability that those conditions are satisfied.

[4] If $\bar{\mathbf{x}}_n = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$, $E(\bar{\mathbf{x}}_n) = E(\mathbf{x})$, $E(\bar{\mathbf{x}}_n^2) = \frac{1}{n} E(\mathbf{x}^2)$. Then if λ is any positive number $P\{|\, \bar{\mathbf{x}}_n - E(\mathbf{x})\,| > \lambda\} \leqq \frac{E\{[\mathbf{x} - E(\mathbf{x})]^2\}}{n\lambda^2}$ which implies (6).

There is also another kind of convergence, called convergence with probability 1. The sequence $\{\mathbf{x}_n\}$ converges with probability 1 to $\mathbf{x}$ if

$$\lim_{n\to\infty} P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda, |\mathbf{x}_{n+1} - \mathbf{x}| \leqq \lambda, \cdots, |\mathbf{x}_{n+p} - \mathbf{x}| \leqq \lambda\} = 1 \tag{7}$$

for every value of $p \geqq 0$, uniformly in $p \geqq 0$ for every positive number λ. If $p = 0$ in (7), (7) becomes (6), so that convergence with probability 1 implies stochastic convergence. Although the converse is not true, if $\{\mathbf{x}_n\}$ is a sequence of chance variables converging stochastically to $\mathbf{x}$, there is a subsequence of $\{\mathbf{x}_n\}$ which converges with probability 1 to $\mathbf{x}$.[5] The usual limit theorems hold here also: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge with probability 1 to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges with probability 1 to $\mathbf{x} + \mathbf{y}$, etc.

An example of convergence with probability 1 is the following. If in the previous example the hypothesis that $E(\mathbf{x}^2)$ exists is removed, so that only the weaker hypothesis of the existence of $E(\mathbf{x})$ is supposed, the Tchebycheff inequality can no longer be applied, but a different method shows that $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ converges with probability 1 (and therefore stochastically) to $E(x)$.[6] This result is known as the strong law of large numbers.

IV

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables with distribution functions $F(x), F_1(x), F_2(x), \cdots$ respectively. Then if $\lim_{n\to\infty} F_n(x) = F(x)$ for every value of x, the distribution of $\mathbf{x}_n$ is said to converge to a limiting distribution with distribution function $F(x)$.

As an example, consider the Laplace-Liapounoff theorem. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of independent chance variables (i.e. any finite number of them form an independent set) with the same distribution functions, and let $E(\mathbf{x}_n)$, $E(\mathbf{x}_n^2)$ exist. We suppose that $\sigma^2 = E\{[\mathbf{x}_n - E(\mathbf{x}_n)]^2\} > 0$ so that the distribution of $\mathbf{x}_n$ is not merely confined to one point. Then the distribution of

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} [\mathbf{x}_j - E(\mathbf{x}_j)] \tag{8}$$

---

[5] The theories of probability and of measure are fundamentally identical. Chance variables correspond to measurable functions. Stochastic convergence corresponds to convergence in measure, and convergence with probability 1 corresponds to convergence almost everywhere. The relation between these two types of convergence is discussed (in the terminology of the measure theory) in E. W. Hobson, The Theory of Functions of a Real Variable, second edition Vol. 2, pp. 239–244.

[6] Cf. for instance J. L. Doob, Transactions of the American Mathematical Society, Vol. 36 (1934), pp. 764–765.

converges to a limiting distribution with distribution function[7]

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2\sigma^2}}\,dx\,. \tag{9}$$

The convergence of a sequence of n-variate distributions is defined as the convergence of the distribution functions just as above for $n = 1$. Suppose that $(\mathbf{x}_{11}, \cdots, \mathbf{x}_{n1})$, $(\mathbf{x}_{12}, \cdots, \mathbf{x}_{n2})$, $\cdots$ are independent sets of chance variables (i.e. the distribution function of any finite number of sets is the product of the distribution functions of the sets) with the same distribution functions. We suppose that $E(\mathbf{x}_{j1})$, $E(\mathbf{x}_{j1}^2)$ exist, $j = 1, \cdots, n$ and that $\sigma_j^2 = E\{[\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]^2\} > 0$. Then if $\bar{\mathbf{x}}_{jm} = m^{-\frac{1}{2}}\sum_{i=1}^{m}[\mathbf{x}_{ji} - E(\mathbf{x}_{ji})]$, the n-variate distribution of $\bar{\mathbf{x}}_{1m}, \cdots, \bar{\mathbf{x}}_{nm}$ converges to the normal distribution[8] about zero means with variances $\sigma_1^2, \cdots, \sigma_n^2$ and correlation coefficients $\{\rho_{ij}\}$ where $\sigma_i\sigma_j\rho_{ij} = E\{[\mathbf{x}_{i1} - E(\mathbf{x}_{i1})]\,[\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]\}$.

Three lemmas will be needed below in applying these concepts.

LEMMA 1. *If $\{\mathbf{x}_n\}$ is a sequence of chance variables whose distributions approach a limiting distribution and if $\{\mathbf{y}_n\}$ is a sequence of chance variables converging stochastically to* 0, *the sequence $\{\mathbf{x}_n\mathbf{y}_n\}$ converges stochastically to* 0.

For if $F(x)$ is the distribution function of the limiting distribution, and if λ, μ are any positive numbers,

$$\begin{aligned} P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} &\geqq P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda,\ \ |\mathbf{y}_n| \leqq \mu\} \geqq P\{|\mathbf{x}_n| < \lambda/\mu,\ \ |\mathbf{y}_n| \leqq \mu\} \\ &\geqq P\{|\mathbf{y}_n| \leqq \mu\} - P\{|\mathbf{x}_n| \geqq \lambda/\mu\} = -P\{|\mathbf{y}_n| > \mu\} + P\{|\mathbf{x}_n| < \lambda/\mu\} \\ &\geqq -P\{|y_n| > \mu\} + P\{\mathbf{x}_n < \lambda/\mu\} - P\{\mathbf{x}_n < -\lambda/2\mu\}\,. \end{aligned} \tag{10}$$

Then, letting n become infinite,

$$\liminf_{n\to\infty} P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} \geqq F(\lambda/\mu) - F(-\lambda/2\mu)\,.^{9} \tag{11}$$

Letting μ approach 0, $F(\lambda/\mu)$ approaches 1, $F(-\lambda/2\mu)$ approaches 0, and the right hand side becomes 1, as was to be proved.

LEMMA 2. *Let $\{\mathbf{x}_n\}$, $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ be sequences of chance variables such that the distribution of x_n approaches a limiting distribution with continuous distribution function $F(x)$ and such that the sequences $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ converge stochastically to* 0, 1 *respectively. Then the distributions of $\{\mathbf{x}_n/\mathbf{z}_n\}$*[10] *and of $\mathbf{x}_n + \mathbf{y}_n$ approach limiting distributions with the same distribution function $F(x)$.*

---

[7] A. Khintchine, Ergebnisse der Mathematik, Vol. 2, No. 4: Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, pp. 1–8.

[8] Ibid. pp. 11–16.

[9] If $\{a_n\}$ is a sequence of real numbers $\limsup_{n\to\infty} a_n$ is defined as $\lim_{n\to\infty}\{\text{least upper bound } a_n, a_{n+1}, \cdots\}$, and $\liminf_{n\to\infty} a_n$ is defined as $-\limsup_{n\to\infty}(-a_n)$. A necessary and sufficient condition that the sequence $\{a_n\}$ converge to a limit a is that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a$.

[10] Since $\mathbf{z}_n$ converges stochastically to 1, the probability that $\mathbf{z}_n = 0$ approaches 0. The theorem is independent of the way $\mathbf{x}_n/\mathbf{z}_n$ is defined when $\mathbf{z}_n = 0$.

Since $\frac{\mathbf{x}_n}{\mathbf{z}_n} = \mathbf{x}_n + \mathbf{x}_n \frac{1 - \mathbf{z}_n}{\mathbf{z}_n}$ (neglecting the possibility that $\mathbf{z}_n$ may vanish), where the last term converges stochastically to 0 by Lemma 1, it is sufficient to prove the second part of the theorem. If $\epsilon > 0$, and if x is an arbitrary number,

$$P\{\mathbf{x}_n + \mathbf{y}_n < x\} = P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} + P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| > \epsilon\}. \tag{12}$$

Since the sequence $\{\mathbf{y}_n\}$ converges stochastically to 0,

$$\lim_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x, |y_n| > \epsilon\} \leqq \lim_{n\to\infty} P\{|\mathbf{y}_n| > \epsilon\} = 0 \tag{13}$$

so that in the limit the second term in (12) can be neglected. Moreover

$$P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} \leqq P\{\mathbf{x}_n < x + \epsilon\}. \tag{14}$$

If we let n become infinite and then let ϵ approach 0, (14) becomes

$$\limsup_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \leqq F(x). \tag{15}$$

A similar argument shows that

$$\liminf_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \geqq F(x), \tag{16}$$

and (15), (16) taken together imply that

$$\lim_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} = F(x), \tag{17}$$

as was to be proved.

LEMMA 3. *If $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ are chance variables whose distribution has density function*

$$\frac{1}{(2\pi)^2} e^{-\frac{1}{2}(x_1^2+x_2^2+x_3^2+x_4^2)}$$

the distribution of $\mathbf{z} = \mathbf{x}_1\mathbf{x}_2 - \mathbf{x}_3\mathbf{x}_4$ has density function $\frac{1}{2}e^{-|x|}$.

The distribution of $\mathbf{u} = \mathbf{x}_1\mathbf{x}_2$ and that of $\mathbf{v} = -\mathbf{x}_3\mathbf{x}_4$ have the same density function:

$$\frac{1}{\pi} \int_0^\infty e^{-\frac{x^2}{2t^2} - \frac{t^2}{2}} \frac{dt}{t}. \tag{18}$$

Hence the distribution of $\mathbf{z}$ has density function

$$\frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_0^\infty \int_0^\infty e^{-\frac{(x-\lambda)^2}{2t^2} - \frac{t^2}{2} - \frac{\lambda^2}{2\tau^2} - \frac{\tau^2}{2}} d\lambda \frac{dt}{t} \frac{d\tau}{\tau}. \tag{19}$$

If we change to polar coördinates: $t = r\cos\theta$, $\tau = r\sin\theta$, and integrate out λ, we obtain

$$\frac{1}{\pi}\int_0^\infty\int_0^{\pi/2} e^{-\frac{x^2}{2r^2}-\frac{r^2}{2}}\,dr d\theta = \frac{1}{2}e^{-|x|}.$$

V

THEOREM 1. *Let* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ *determine a 4-variate distribution with distribution function* $F(x_1, x_2, x_3, x_4)$. *Suppose that* $E(\mathbf{x}_i)$, $E(\mathbf{x}_i^2)$, $E(\mathbf{x}_i^2\mathbf{x}_j^2)$ *exist,* $i, j = 1, \cdots, 4$, *and suppose that* $E(\mathbf{x}_i) = 0$, $E(\mathbf{x}_i^2) = 1$,[11] $i, j = 1, 2, 3, 4$. *Let* $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \mathbf{x}_{3j}, \mathbf{x}_{4j}$ *have the same 4-variate distribution as* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$, $j = 1, \cdots, n$, *and let the 4n-variate distribution function of* $\{\mathbf{x}_{ij}\}$ *be* $\prod_{j=1}^{n} F(x_{1j}, x_{2j}, x_{3j}, x_{4j})$. *We shall use the following notation (which suppresses the dependence on n):*

$$\xi_i = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}, \qquad \mathbf{s}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}\mathbf{x}_{jk}, \qquad \rho_{ij} = E(\mathbf{x}_i\mathbf{x}_j). \tag{20}$$

Let φ *be a function of* ξ_i, s_{ij}, *defined in a neighborhood N of P:* $\xi_i = 0$, $s_{ij} = \rho_{ij}$, *which, together with its second partial derivatives is continuous in N. Define* $\sigma \geqq 0$ *by*

$$\sigma^2 = E\left\{\left[\sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\mathbf{x}_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - \mathbf{x}_i\mathbf{x}_j)\right]^2\right\}, \tag{21}$$

where the partial derivatives are evaluated at P. Then if $\sigma > 0$, *the distribution of* $\sqrt{n}\,[\varphi - \varphi(P)]$ (*where* φ *has the arguments* ξ_i, $\mathbf{s}_{ij}$) *converges to a limiting distribution which is normal with mean* 0 *and variance* σ^2.

To prove this theorem we expand φ in the neighborhood of P, obtaining

$$\sqrt{n}\,[\varphi - \varphi(P)] = \sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\sqrt{n}\,\xi_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij}) + \mathbf{R}_n \tag{22}$$

where the partial derivatives are evaluated at P, and where $\mathbf{R}_n$ consists of a linear combination of $\sqrt{n}\,\xi_i\xi_j$, $\sqrt{n}\,\xi_i(\rho_{jk} - \mathbf{s}_{jk})$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})\,(\rho_{kl} - \mathbf{s}_{kl})$, with coefficients which are uniformly bounded as long as ξ_i, $\mathbf{s}_{ij}$ are in the neighborhood N. Now

$$\lim_{n\to\infty}\xi_i = 0 \qquad \lim_{n\to\infty}\mathbf{s}_{ij} = \rho_{ij} \tag{23}$$

with probability 1, by the law of large numbers, and as n becomes infinite the distributions of $\sqrt{n}\,\xi_i$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})$ converge to limiting distributions, by the

[11] The hypothesis that $E(\mathbf{x}_i) = 0$ involves no real restriction, since the general case can be reduced to this one by substituting $\mathbf{x}_i - E(\mathbf{x}_i)$ for $\mathbf{x}_i$. The hypothesis that $E(\mathbf{x}_i^2) = 1$ can be met by substituting $\mathbf{x}_i[E(\mathbf{x}_i^2)]^{-\frac{1}{2}}$ whenever $E(\mathbf{x}_i^2) > 0$, which will always be true unless $\mathbf{x}_i = 0$ with probability 1.

Laplace-Liapounoff theorem. Then by Lemma 1, the terms of $\mathbf{R}_n$ converge stochastically to 0. The other terms of $\sqrt{n}[\varphi - \varphi(P)]$ are sums to which the Laplace-Liapounoff theorem can be applied, giving the desired conclusion.

As an example of the application of this theorem, we suppose that φ is a correlation coefficient:

$$\varphi = \frac{\mathbf{S}_{12}}{(\mathbf{S}_{11}\mathbf{S}_{22})^{\frac{1}{2}}}, \qquad \varphi(P) = \rho_{12}. \tag{24}$$

Here σ^2 is $E\{[\mathbf{x}_1\mathbf{x}_2 - \frac{1}{2}\rho_{12}(\mathbf{x}_1^2 + \mathbf{x}_2^2)]^2\}$, (which reduces to the familiar result $1 - \rho_{12}^2$ when the bivariate distribution of $\mathbf{x}_1$, $\mathbf{x}_2$ is normal) and $\sigma = 0$ only when, with probability 1,

$$2\,\mathbf{x}_1\mathbf{x}_2 = \rho_{12}(\mathbf{x}_1^2 + \mathbf{x}_2^2). \tag{25}$$

As a second example we suppose that φ is a tetrad difference:

$$\varphi = \frac{\mathbf{S}_{13}\mathbf{S}_{24} - \mathbf{S}_{14}\mathbf{S}_{23}}{(\mathbf{S}_{11}\mathbf{S}_{22}\mathbf{S}_{33}\mathbf{S}_{44})^{\frac{1}{4}}}, \qquad \varphi(P) = \rho_{13}\rho_{24} - \rho_{14}\rho_{23}. \tag{26}$$

Here σ^2 becomes

$$\sigma^2 = E\left\{\left[\rho_{24}\mathbf{x}_1\mathbf{x}_3 + \rho_{13}\mathbf{x}_2\mathbf{x}_4 - \rho_{14}\mathbf{x}_2\mathbf{x}_3 - \rho_{23}\mathbf{x}_1\mathbf{x}_4 - \frac{\varphi(P)}{2}\sum_{j=1}^{4}\mathbf{x}_j^2\right]^2\right\} \tag{27}$$

and $\sigma = 0$ only when the quantity in the brackets vanishes with probability 1.

If in either of the two above cases $s_{ij} - \xi_i\xi_j$ is substituted for s_{ij} (i.e. if the deviations from the sample mean, not those from the true mean, are used), the result is unaltered. This is true in general, since $\frac{\partial\varphi}{\partial\xi_i}$, $\frac{\partial\varphi}{\partial s_{ij}}$ are unaltered at P by this substitution.

There is a well-known δ-method used in statistics to find limiting variances of statistics of the type covered by Theorem 1,[12] and Theorem 1 shows an interpretation which can be given to the results obtained by this method.

We now investigate the necessary modification of Theorem 1 if $\sigma = 0$, i.e. if

$$\sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\mathbf{x}_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - \mathbf{x}_i\mathbf{x}_j) = 0 \tag{28}$$

with probability 1. If we assume that φ has continuous third partial derivatives in the neighborhood N, we find that

[12] Examples of the use of this method can be found in T. L. Kelley, Crossroads in The Mind of Man, Stanford University (1928), pp. 49–50, and in an article by S. Wright, Annals of Mathematical Statistics, Vol. 5 (1934), p. 211.

$$n[\varphi - \varphi(P)] = \frac{n}{2}\sum \frac{\partial^2\varphi}{\partial\xi_i\xi_j}\,\xi_i\xi_j + \frac{n}{2}\sum_{i,j,k}\frac{\partial^2\varphi}{\partial\xi_i\,\partial s_{jk}}\,\xi_i(\mathbf{s}_{jk} - \rho_{jk}) + \frac{n}{2}\sum_{i,j,k,l}\frac{\partial^2\varphi}{\partial s_{ij}\,\partial s_{kl}}\,(\mathbf{s}_{ij} - \rho_{ij})(\mathbf{s}_{kl} - \rho_{kl}) + \mathbf{R}_n' \tag{29}$$

where $\mathbf{R}_n'$ converges stochastically to 0. The second degree terms constitute a quadratic form in $\{\xi_i,\ \mathbf{s}_{jk} - \rho_{jk}\}$. Now the multivariate distribution of $\{\sqrt{n}\xi_i,\ \sqrt{n}(\mathbf{s}_{jk} - \rho_{jk})\}$, by the Laplace-Liapounoff theorem, converges to a normal distribution whose variances and correlation coefficients are those of $\mathbf{x}_i,\ \mathbf{x}_j\mathbf{x}_k$. The distribution of $n[\varphi - \varphi(P)]$ thus converges to the distribution of the quadratic form

$$\frac{n}{2}\sum_{i,j}\frac{\partial^2\varphi}{\partial\xi_i\,\partial\xi_j}\,\mathfrak{a}_i\mathfrak{a}_j + \frac{n}{2}\sum_{i,j,k}\frac{\partial^2\varphi}{\partial\xi_i\,\partial s_{jk}}\,\mathfrak{a}_i\mathfrak{b}_{jk} + \frac{n}{2}\sum_{i,j,k,l}\frac{\partial^2\varphi}{\partial s_{ij}\,\partial s_{kl}}\,\mathfrak{b}_{ij}\mathfrak{b}_{kl}, \tag{30}$$

where $\{\alpha_i,\ \beta_{jk}\}$ have the multivariate distribution just described, unless the quadratic form vanishes identically. This reasoning can be continued, the general result being that there is some power ν of n, if φ is sufficiently regular, such that the distribution of $n^\nu[\varphi - \varphi(P)]$ converges to a limiting distribution.

When $\sigma = 0$ in the second example, unless the distribution of $\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_4$ is confined with probability 1 to a 4-dimensional quadric, $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$. Equation (29) becomes

$$n[\varphi - \varphi(P)] = \mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23} + \mathbf{R}_n'. \tag{29'}$$

Now if $\mathbf{x}_1,\ \mathbf{x}_2$ are transformed by a linear homogeneous transformation with determinant Δ, it is readily seen that $\mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23}$ is multiplied by Δ. The same is true of $\mathbf{x}_3,\ \mathbf{x}_4$. If $\mathbf{x}_1,\ \mathbf{x}_2$ are transformed into $\mathbf{x}_1',\ \mathbf{x}_2'$ so that $E(x_j'^2) = 1$, $E(x_1'x_2') = 0$, the determinant of the transformation is $\pm(1 - \rho_{12}^2)^{-\frac{1}{2}}$. Then transforming each pair $(\mathbf{x}_1,\ \mathbf{x}_2)$, $(\mathbf{x}_3,\ \mathbf{x}_4)$ in this way into $(\mathbf{x}_1',\ \mathbf{x}_2')$, $(\mathbf{x}_3',\ \mathbf{x}_4')$, the variables $\mathbf{x}_1',\ \mathbf{x}_2',\ \mathbf{x}_3',\ \mathbf{x}_4'$ are uncorrelated. If $\mathbf{s}_{ij}' = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}'\mathbf{x}_{jk}'$,

$$\mathbf{s}_{13}'\mathbf{s}_{24}' - \mathbf{s}_{14}'\mathbf{s}_{23}' = \frac{\mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23}}{\pm(1 - \rho_{12}^2)^{\frac{1}{2}}(1 - \rho_{34}^2)^{\frac{1}{2}}}. \tag{31}$$

The limiting distribution of $\mathbf{s}_{13}'\mathbf{s}_{24}' - \mathbf{s}_{14}'\mathbf{s}_{23}'$ is the distribution of $\mathfrak{b}_{13}'\mathfrak{b}_{24}' - \mathfrak{b}_{14}'\mathfrak{b}_{23}'$ where these four chance variables are normally distributed, $E(\mathfrak{b}_{13}') = E(\mathfrak{b}_{24}') = E(\mathfrak{b}_{14}') = E(\mathfrak{b}_{23}') = 0$, $E(\mathfrak{b}_{ij}') = E(\mathbf{x}_i'\mathbf{x}_j')$, $E(\mathfrak{b}_{ij}'\mathfrak{b}_{kl}') = E(\mathbf{x}_i'\mathbf{x}_j'\mathbf{x}_k'\mathbf{x}_l')$. Now if $\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_4$ are normally distributed—the most important case for statistical purposes—$\mathbf{x}_1',\ \mathbf{x}_2',\ \mathbf{x}_3',\ \mathbf{x}_4'$ will also be distributed normally, and the vanishing of the correlation coefficients means that the chance variables are independent. If this is true

$$E(\mathfrak{b}_{ij}'^2) = 1, \qquad E(\mathfrak{b}_{ij}\mathfrak{b}_{kl}) = 0, \qquad (\mathfrak{b}_{ij} \neq \mathfrak{b}_{kl}). \tag{32}$$

Evidently, however, $\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3', \mathbf{x}_4'$ do not have to be independent to make these equations valid. It is more than sufficient if the pairs $(\mathbf{x}_1, \mathbf{x}_2)$, $(\mathbf{x}_3, \mathbf{x}_4)$ and therefore the pairs $(\mathbf{x}_1', \mathbf{x}_2')$, $(\mathbf{x}_3', \mathbf{x}_4')$ are independent. If (32) is true, the β's are independent, each one being normally distributed with mean 0 and variance 1. Summarizing these results, and using Lemma 3: *if φ is the tetrad difference and if $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, the distribution of $n[\varphi - \varphi(P)]$ converges to a limiting distribution. If in addition the distribution of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ is normal, or if the pairs $(\mathbf{x}_1, \mathbf{x}_2)$ $(\mathbf{x}_3, \mathbf{x}_4)$ are independent, this limiting distribution has density function*

$$\frac{c}{2} e^{-c|x|}$$

where $c = (1 - \rho_{12}^2)^{-\frac{1}{2}} (1 - \rho_{34}^2)^{-\frac{1}{2}}$.

Wilks has investigated the case where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ are normally and independently distributed, and in this case found the exact variance of the tetrad difference as a function of n.[13]

COLUMBIA UNIVERSITY.

[13] Proceedings of the National Academy of Sciences, Vol. 18, (1932), pp. 562–565.